How a long-running rainforest study nurtured Peruvian science p. 508

Eppendorf & *Science* Prize for how gut glial cells feel the force p. 527

Blazing fast thermal switching p. 585

Science

$15
3 NOVEMBER 2023
science.org

AAAS

HEAVY HERBIVORY

Plant consumption limits restoration success

pp. 516 & 589

EDITORIAL

Biodiversity market doublespeak

Michael J. Vardon is an associate professor at the Fenner School of Environment and Society, Australian National University, Canberra, Australia. michael.vardon@anu.edu.au

David B. Lindenmayer is a professor at the Fenner School of Environment and Society, Australian National University, Canberra, Australia. david.lindenmayer@anu.edu.au

The notion that biodiversity markets can raise money desperately needed for biodiversity conservation is gaining momentum. The dire state of biodiversity and the enormous biodiversity repair bill means that every funding option must be explored. However, the risk that trading ill-defined generic biodiversity credits will result in biodiversity loss, not conservation, should be considered. Scarce resources could be diverted to market regulation rather than conservation. Without key elements, biodiversity markets could be perverse, leading to Orwellian "doublespeak"—saying one thing, but resulting in another.

Target 19d of the 2022 Kunming-Montreal Global Biodiversity Framework features private-sector conservation finance, and various businesses, international agencies, and countries are embracing biodiversity markets. Such efforts recognize that many actors are required by law, or wish voluntarily, to support ecosystem conservation or restoration, but may not be well-positioned to do so directly and seek to satisfy their needs through biodiversity markets. The logic is that these markets are necessary for conservation to be an economically efficient and viable alternative to current land uses.

> "Without key elements, biodiversity markets could be perverse..."

For example, a Nature Repair Market Bill is before Australia's parliament to establish a voluntary biodiversity market. The market is in addition to the existing biodiversity offsets needed for regulatory compliance. It remains to be seen if the Bill passes into law and can deliver the promised environmental-economic "win-win" and not become conservation doublespeak.

A critical challenge is the private sector's desire for universal "biodiversity credits." However, combining heterogenous biodiversity into a universal unit is difficult. Adding two black flying foxes (*Pteropus alecto*) to two Leadbeater's possums (*Gymnobelideus leadbeateri*) does not equal four "biodiversities." If this sounds silly, it is. A "common currency" for biodiversity is a long-known problem in biodiversity offset markets. No universal offset exists, and simplistic metrics usually fail to deliver expected outcomes. Homogenized biodiversity credits should thus be dismissed, but species- or habitat-specific biodiversity credits may provide an answer.

Biodiversity market enthusiasts recognize the need for safeguards: "Integrity and strong governance are non-negotiable," but little detail usually exists. Yet detail is essential as biodiversity markets are created for biodiversity conservation, not economic gain. Laws already govern markets for goods and services, with many agencies safeguarding market participants and minimizing unwanted socioeconomic and environmental impacts. All of this must apply to biodiversity trading as well.

Laws establishing biodiversity markets would not supplant existing conservation laws. New laws should determine what is traded (e.g., species-specific biodiversity credits), how credits are calculated, reporting standards, and the rights and obligations of credit owners and the owners or managers of the underlying biodiversity. The last is critical to ensure accountability for conservation.

For markets to achieve biodiversity conservation, biodiversity must endure. A "buy and forget" model, where credits exist forever regardless of whether the biodiversity persists, would lead to perverse outcomes. To prevent this, monitoring, reporting, and enforcement are critical. Traditional accounting records the volume and value of biodiversity credit trades and holdings. Biodiversity trading also requires accounting for biodiversity itself: how much there is, how and why it is changing, and what is being spent on management. The accounting must link credits to the species underpinning credits, which requires systematic and ongoing monitoring; Leadbeater's possum credits need a clear relationship to Leadbeater's possums. The Task Force on Nature-related Financial Disclosures and System of Environmental Economic Accounting provide accounting guidance, but the accounting must be standardized and mandated, not optional.

An independent agency with biodiversity and financial expertise—for example, a Natural Capital Bank—would monitor and analyze an organization's biodiversity accounts. Whatever the agency, the accounts provide an audit trail for compliance and enforcement, and it must work with existing organizations responsible for market regulation and environmental protection. Funding should come from biodiversity trading (e.g., a 2.5% levy on every trade), not existing conservation resources.

The push for biodiversity markets and private-sector conservation finance is understandable. But unless biodiversity markets are well designed, there is a risk they will become conservation doublespeak, legitimizing biodiversity destruction for economic gain while purporting to promote biodiversity conservation.

–Michael J. Vardon* and David B. Lindenmayer

*M.J.V. has been an adviser on the System of Environmental-Accounting to the World Bank, United Nations, and projects sponsored by the governments of Australia, the Netherlands, New South Wales, and the Australian Capital Territory.

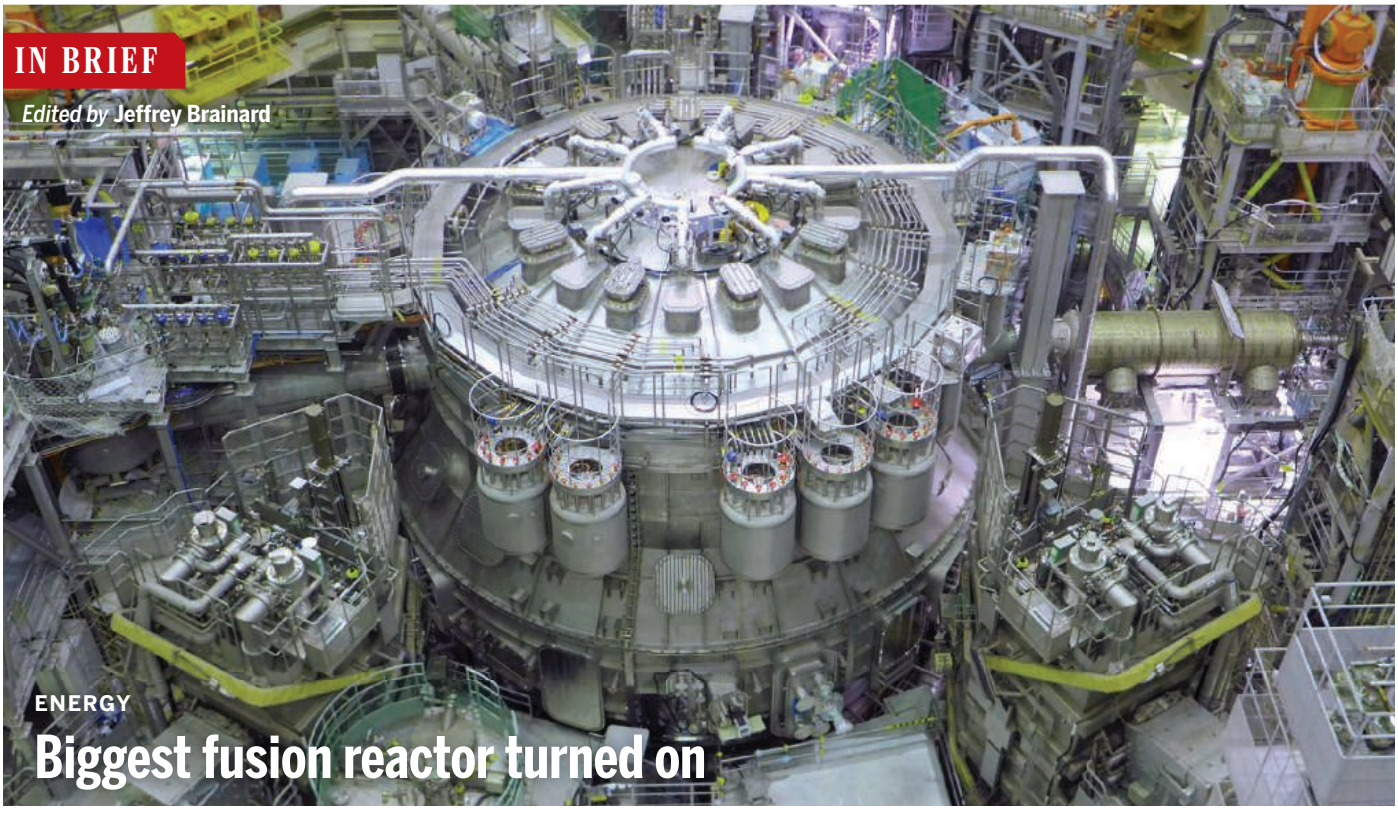
10.1126/science.adg6823

NEWS

162% Projected increase by midcentury in annual excess U.S. cardiovascular deaths associated with extreme heat, if climate change warms Earth's temperature by 2°C. The excess deaths would total 4320 a year. (*Circulation*)

IN BRIEF

Edited by **Jeffrey Brainard**



ENERGY

Biggest fusion reactor turned on

Operators have fired up the world's newest and largest experimental fusion reactor, designed as a stepping stone to commercial devices that could yield carbon-free electricity. Workers began to construct the JT-60SA reactor (above)—a joint project of the European Union and Japan, which hosts the device—in 2013. The doughnut-shaped reactor, 12 meters in diameter, reached a milestone last week when it used magnetic fields generated by superconducting coils to contain a cloud of ionized gas, or plasma. Eventually the device will contain plasma heated to 200 million degrees Celsius for about 100 seconds, far longer than previous large tokamaks. JT-60SA will test technologies to be used in ITER, the international fusion reactor being built in France to demonstrate that fusion can generate more energy than goes into producing it. ITER's plasma volume will be almost six times JT-60SA's.

Fossil fuel use could peak by 2030

ENERGY | Global consumption of oil, natural gas, and coal is likely to peak by 2030 as the world switches to renewable sources, the International Energy Agency said last week. But developing nations will continue to increase burning of fossil fuels after 2030 as they rapidly industrialize, according to the agency's annual forecast. To reach net zero emissions globally by 2050, those nations will need to expand their spending on construction of clean-energy sources to nearly $2 trillion annually, seven times the current level. Their expenditures on clean energy have stagnated since 2015 and increasing them will probably require loans from wealthy countries on favorable terms, the report says.

Gene therapy restores hearing

BIOMEDICINE | Several children born deaf can now hear after receiving gene therapy. The children, from China, had inherited two defective copies of the gene encoding otoferlin, a protein that helps the inner ear transmit sound to the brain. A team at Fudan University injected harmless viruses carrying DNA for a working copy of the hearing gene into one inner ear of each child. Four of five kids treated now have some hearing, the team reported last week at the annual meeting of the European Society for Gene and Cell Therapy in Belgium, according to the group's conference abstract and *MIT Technology Review*. The U.S. company Regeneron last week reported similar

success for the first child treated with its comparable gene therapy. Deafness due to mutations in otorferlin's gene is rare—it explains up to 8% of cases of inherited deafness—but the results offer hope for treating other genetic forms of deafness.

U.K. funder halts diversity panel

POLITICS | The U.K.'s national science funding agency has suspended operations of its newly formed diversity advisory panel after science minister Michelle Donelan expressed "outrage" at social media posts about the Israel-Hamas conflict by two of its 11 members. In a 28 October letter to UK Research and Innovation (UKRI) head Ottoline Leyser, Donelan asked the funding agency to disband the panel on equality, diversity, and inclusion. More than 2500 researchers signed an open letter calling on UKRI to reject Donelan's demands. But in a statement published on 31 October, Leyser said UKRI would suspend the panel and launch an investigation to determine "the ultimate future" of the board. The University and College Union, which represents higher education workers, said Donelan's letter was "unwarranted pressure … by an increasingly authoritarian government" and that UKRI's move "set a very dangerous precedent for academic freedom."

Disgraced M.D. inspires opera

BIOMEDICINE | "Who has the courage to unmask a liar who everyone wants to believe?" asks the ad for *Mythomania*, a new opera based on the saga surrounding disgraced surgeon Paolo Macchiarini. He claimed to have developed a way to regenerate a damaged or diseased windpipe using stem cells, but the technique did not work and almost all the patients he treated died from related complications. He is appealing a 30-month prison sentence he received from a Swedish court this year for gross assault against three of the patients who died after he treated them at the Karolinska Institute. Last week, *The Lancet* retracted papers he published in 2008 and 2013 about the technique. The opera's composer, Paula af Malmborg Ward, and librettist Kerstin Perski created a previous opera about swindlers and wanted to further explore why people believe serial liars, they told the Swedish press. A group from the Swedish National Board for Assessment of Research Misconduct, which investigated Macchiarini, was expected to attend the opera, which is playing in Gothenburg, Sweden, through 9 November.

PHOTO: KEVIN SCHAFER/ALAMY

A brown skua nests above a penguin colony on Bird Island.

CONSERVATION

Lethal bird flu approaches Antarctica

A virulent strain of avian influenza has been confirmed for the first time in the Antarctic region. Scientists with the British Antarctic Survey noticed dead brown skuas on Bird Island in South Georgia and the South Sandwich Islands starting on 7 October. The virus, H5N1, was likely carried to the island by migratory birds returning from South America, which has been hit hard. South Georgia's government has banned visitors to the small island, which is less than 5 kilometers long, but research on the bird and seal colonies will continue. "Bird Island is one of the most closely monitored seabird and seal colonies in the world, so ongoing studies will reveal the impacts of the disease in detail," the government said in a statement. Scientists are also monitoring to detect the virus' expected move to the Antarctic continent, home to much larger wildlife populations.

Muscular dystrophy therapy falters

BIOMEDICINE | A much-anticipated clinical trial of a gene therapy failed to clearly show that it helps young children developing symptoms of Duchene muscular dystrophy (DMD). In the placebo-controlled, phase 3 trial of 125 DMD patients, ages 4 to 7, the intervention did not meet its primary endpoint, a statistically significant preservation of motor function, Sarepta Therapeutics, the gene therapy's developer, announced this week. Earlier this year, the U.S. Food and Drug Administration gave the treatment—in which viruses deliver a gene for a short form of the muscle protein dystrophin into muscle tissue—a tentative "accelerated approval" limited to 4- and 5-year-olds possessing a dystrophin gene mutation. Sarepta defended the new trial's results, saying treated children did modestly better than the placebo arm on the motor function measure and also better retained specific abilities, such as walking, running, and rising from the floor. The company still intends to seek a wider approval of the one-dose gene therapy, for which it is charging more than $3 million.

Russia drops nuclear test treaty

ARMS CONTROL | Russia's parliament last week fulfilled a request by President Vladimir Putin to revoke the country's ratification of the international Comprehensive Nuclear-Test-Ban Treaty, heightening concerns that it plans to resume testing nuclear bombs. Russia will continue to host instruments in the treaty's monitoring network to detect clandestine nuclear tests, officials say. Since the United Nations adopted it in 1996, the treaty has established norms that have limited almost all explosive nuclear testing. Leaders of 187 nations have signed the treaty, but it has not taken effect because eight nuclear-capable states, including the United States, have not ratified it. Russia says it wishes to achieve "strategic parity" with the U.S. Weapons labs in Russia, China, and the U.S. have reportedly been increasing bomb-design activities. Analysts say Russia's action may be a response to Western military aid for Ukraine.

IN DEPTH

ENDANGERED SPECIES

Lawsuits force changes in U.S. pesticide regs

EPA is drawing up new rules for farmers, but some worry the science isn't ready

By **Erik Stokstad**

Driven by tight court deadlines, the U.S. Environmental Protection Agency (EPA) is radically remaking its approach to regulating pesticides, giving weight to their risk to endangered species. The effort is still in the early stages, but concrete enough that it has alarmed agricultural organizations, which fear widespread restrictions on where and how farmers can spray the chemicals. "It can potentially add really tremendous cost," says James Cranney, head of the California Citrus Quality Council.

Before approving pesticides for use, EPA evaluates the risk they pose to human health and the environment. As part of those analyses, the agency is legally required to consider potential impacts on endangered species. But for decades it has rarely completed the reviews, which require consultation with the U.S. Fish and Wildlife Service (FWS) and National Marine Fisheries Service (NMFS) and often drag out for more than 4 years.

In a draft plan that was open for comment until last month, the agency proposed to speed up the process by making its own internal assessment of risk to endangered species. EPA will continue to consult with FWS and NMFS, as the law requires. But it won't wait for the other agencies to finish weighing in before requiring farmers to change their practices. To add efficiency, the agency will also conduct analyses looking at the risks pesticides pose to groups of organisms.

Environmental groups hope the new regulatory approach will strengthen protections for species such as the rusty patched bumble bee, which used to occur widely over the north-central and eastern United States but for reasons including pesticide use, is now scattered across just a few states. "We're thrilled that EPA is diving into this," says Aimée Code, who heads the pesticide program at the Xerces Society for Invertebrate Conservation.

Others caution that the pace is too fast. "Decisions are being made before the science is ready," says Stanley Culpepper, a weed scientist at the University of Georgia who would like to see more detailed research on where endangered species are present and their sensitivity to various pesticides. And the potential economic impacts worry not just farm groups, but also the U.S. Department of Agriculture, which suggested in a comment to EPA that the new regulations could drive farmers in some areas to sell off their land for development.

EPA's new strategy was kick-started by lawsuits, including a 2011 case brought by the Center for Biological Diversity and the Pesticide Action Network North America. The groups claimed EPA had violated the law by not considering endangered species over the course of approving more than 1000 pesticide products. In September, EPA agreed to a final settlement stipulating that it would assess the risks of harm from 12 active ingredients in these pesticides, and that it would come up with strategies to evaluate and minimize risks to endangered species from new pesticides as well.

The effort is headed by Jake Li, who was appointed EPA's deputy assistant administrator for pesticide programs in 2021. Trained as a lawyer, Li worked on endangered species policy at the Environmental Policy Innovation Center and at Defenders of Wildlife, where he filed a pesticide lawsuit against the office where he now works. "This is a sprint right now for us after 40 years of neglect," Li says. "We are trying to make the most sense out of a very difficult problem."

EPA is initially focusing on agricultural weedkillers. When assessing risk, it plans to group the more than 900 endangered species that may live in and around farmland according to their biology, such as the microhabitat they live in. Animals and plants that inhabit small wetlands near fields, for example, might be assessed as facing higher odds of harm from agricultural herbicides than those that live on dry ground or far-

The rusty patched bumble bee was put on the U.S. endangered species list in 2017.

ther away. Those findings will then inform the conditions EPA sets for using a product in a particular location.

EPA is also developing more targeted regulations to protect species that are especially vulnerable. In a draft white paper released this summer, the agency presented the results of a pilot project that reviewed data on the susceptibility of 27 species to herbicides, insecticides, and other crop protection chemicals. The draft proposed restrictions on farmers depending on what species are nearby. To protect species such as the rusty patched bumble bee and the American burying beetle, for example, EPA proposes prohibiting spraying with fine droplets in order to lessen drift, and other actions.

Some researchers worry the maps of species ranges in the pilot project weren't detailed enough. For example, the Taylor's checkerspot butterfly—a species that is susceptible to insecticides used to control moth pests—is shown as being present across 2 million hectares of Oregon and Washington state, including prime agricultural land. But in reality, the species is only found in 13 patches of surviving prairie, says Cheryl

Taylor's checkerspot butterflies were once abundant in prairie habitat in the Pacific Northwest.

PHOTO: AARON BARNA/U.S. FISH AND WILDLIFE SERVICE

Schultz, a conservation biologist at Washington State University.

Culpepper argues there's a need for "significant investment" in mapping endangered species. He points to a 2022 EPA decision that restricted the use of the herbicide Enlist Duo in more than 43 counties nationwide. In Georgia, where the restrictions were enacted because of concern for frosted and flatwood salamanders, the decision meant Enlist Duo could not be used on more than 385,000 hectares of cotton, corn, and soybean crops. After the decision, Culpepper worked with FWS and state agencies to more finely map the salamanders' habitat and found it was only necessary to restrict spraying on 1427 hectares.

Something similar happened in Oklahoma and Texas, where cotton farmers briefly couldn't use Enlist Duo last year because of risk to the endangered American burying beetle. "It was a good idea, poorly implemented," says Wyatt Hoback, an entomologist at Oklahoma State University who points out that the beetle isn't found in cotton fields. Hoback is hopeful that EPA's new regulatory approach won't run into similar problems. Excessive regulation could be counterproductive, he says: "It's just going to create more animosity toward rare species." Li emphasizes that the white paper is a draft and will be revised when inaccurate. "We are not going to finalize it unless we can get defensible maps."

Many observers would like to see EPA's modeling become more realistic as well. EPA analysts predict risk to wildlife by taking toxicity data from laboratory studies and then making assumptions about how pesticides are applied on farms. One weakness of the approach for terrestrial insects, entomologists point out, is that pesticides are only required to be tested on domesticated honey bees, so it's not always clear how wild bumble bees and other species will be affected. And agricultural groups note that farmers often spray less than EPA's modeling assumes. Li says EPA would gladly accept data on pesticide usage.

The agency plans to finalize its herbicide strategy in May 2024 and complete its list of mitigation activities for the 27 vulnerable species by the end of this year. "There might be some missteps with this pilot program, but the fact they're getting information on doing it better—that's fabulous," Schultz says. "There's enormous potential."

Then, EPA will move from herbicides to a harder challenge: insecticides. Assessing their risk to endangered species will be more complicated because insecticides can harm wildlife in many ways, including neurological damage that isn't lethal but impairs the ability of pollinators to forage. And EPA does not regulate seeds coated with insecticides, such as neonicotinoids—a major source of exposure to insects. The agency is looking into the issue, Li says, and last month issued a request for public comment about regulation of pesticide-treated seeds.

"I'm cautiously optimistic," Code says of EPA's plans. "There are a lot of growers who want to do right by the species, and I feel like … if we get people behind this, it could really be a game changer." ■

COSMOLOGY

No easy way to explain cosmic expansion mystery

"Hubble tension" could be a signal of new physics. But deciphering it may not be simple

By **Adrian Cho**

Discovered less than a century ago, the expansion of the universe causes galaxies to rush away from Earth, stretching their light to longer, redder wavelengths. That observation spawned the idea of the big bang—and decades of bickering over the rate at which the universe is expanding, the Hubble constant. After a brief rapprochement, cosmologists are arguing again. Working from our cosmic neighborhood outward to more distant galaxies, one group has measured a rate significantly higher than the one derived by colleagues studying the cosmos' farthest fringe and the afterglow of the big bang, the cosmic microwave background (CMB).

Recent observations from NASA's JWST space telescope have confirmed the disparity, or Hubble tension, and chances that it can be explained as an observational artifact are fading. But so are prospects that the puzzle has a simple physics solution, a recent spate of papers shows. That may dash hopes that solving the Hubble tension could also help cosmologists sharpen their problematic theory of the universe's makeup and evolution.

"There's no guarantee that there's one effect that is causing all of this," says Adam Riess, a cosmologist at Johns Hopkins University. Some question whether the Hubble tension will ever be explained. "I wouldn't bet my house on it," says Sunny Vagnozzi, a cosmologist at the University of Trento.

According to cosmologists' prevailing theory, the universe contains 5% ordinary matter; 27% invisible dark matter, whose gravity holds galaxies together; and 68% dark energy, which stretches space like a pressure. Just after the big bang, the universe grew exponentially, hugely magnifying tiny quantum fluctuations in a dense soup of fundamental

particles. Dark matter gathered in the dense spots, and hundreds of millions of years later ordinary matter settled in the clumps of dark matter to form galaxies. As the universe thinned, dark energy's push overcame the pull of gravity, so that after slowing, the universe's expansion is now accelerating.

The model, called Lambda–cold dark matter (Lambda-CDM), is vague. It assumes dark energy is just a property of empty space called a cosmological constant, Lambda, and dark matter is just cold invisible stuff, hence CDM. But the model fits cosmological data beautifully, especially those from the CMB. The temperature of the CMB varies ever so slightly across the sky, providing a blotchy snapshot of the universe at the moment neutral atoms formed, 380,000 years after the big bang. A theory with just six parameters—including the Hubble constant—Lambda-CDM precisely fits the distribution of fluctuation sizes measured by Europe's Planck spacecraft, says Tanvi Karwal, a cosmologist at the University of Chicago. "It's insane that it works."

However, the fit yields a value of the Hubble constant that clashes with the one measured directly. Since 2009, Riess and his colleagues have used various telescopes to create an elaborate "distance ladder" of the distances and red shifts of nearby galaxies. A key rung relies on observations of variable stars called Cepheids. Each pulses at a rate that reveals its intrinsic brightness, allowing observers to deduce its distance from its apparent brightness in the sky. JWST can tease out individual Cepheids in other galaxies, and its observations confirm the Hubble constant is 8% higher than Lambda-CDM predicts, Riess and colleagues reported in September.

Some researchers say they're still not 100% convinced the discrepancy is real. "The field really needs a sensitive third way of measuring this," says Johannes Eskilt, a cosmologist at the University of Oslo. Others have proposed myriad models to explain it.

One solution assumes dark energy isn't a cosmological constant, but is due to some kind of new physics. If so, its concentration could have fallen or even grown as the cosmos evolved, resulting in an expansion history that could start as the Lambda-CDM predicts but end at the higher directly measured value. In fact, Riess says he started his project in hopes of finding just such an effect.

Other data nix that idea. Sound waves called baryon acoustic oscillations (BAOs) rippled through the infant universe, and the CMB records how far they spread before atoms formed. That length is also imprinted in the distribution of galaxies. Comparing the two lengths reveals the expansion history halfway between the big bang and now. The result is a curve whose shape matches Lambda-CDM's prediction, says Ryan Keeley, a cosmologist the University of California, Merced. That constraint makes it impossible to change the expansion history enough to resolve the Hubble tension, he and a colleague argued on 15 September in *Physical Review Letters* (*PRL*). "These most recent BAO data sets put the final nail in the coffin" of that solution, Keeley says.

The answer must lie earlier in cosmic history, says Marc Kamionkowski, a cosmologist at Johns Hopkins. "We're forced to look at the early time solutions and take them more seriously," he says. Those theories boost the expansion rate before the CMB emerged, allowing the cosmos to evolve much as Lambda-CDM predicts while obtaining a higher Hubble constant now. In 2016, Karwal and Kamionkowski proposed that a new quantum field, a bit like an electric field, kicked in briefly 100,000 years after the big bang and acted like extra "early" dark energy. Others have suggested the extra kick could have come from primordial magnetic fields, interactions of neutrinos, or even changes in the electron's mass.

Although more promising, such models also have problems, Vagnozzi says. For example, the CMB is known to get slightly blue shifted as it passes through clumps of dark matter. The effect gets bigger if the youthful universe expands faster. To avoid contradicting observations, early time solutions must include more dark matter. But the extra dark matter would tend to make the mature universe clumpier than it is, Vagnozzi says.

For example, light from distant, brilliant galaxies called quasars reveals no sign of extra clumpiness, says Samuel Goldstein, a cosmologist at Columbia University. A quasar's spectrum shows a thicket of dark absorption lines, known as the Lyman-alpha forest. Each line marks a cloud of hydrogen that lies along the line of sight and soaks up the quasar's light. In a paper in press in *PRL*, Goldstein and colleagues report that the clouds clump less than early dark energy would predict. "We ended up getting ridiculously strong constraints on early dark energy."

The clumpiness issue is one reason why something like early dark energy alone won't solve the Hubble tension, Vagnozzi argues. "Early universe physics is always going to fall short and there will have to be something else," he says. Karwal is less convinced. The universe is already less clumpy than Lambda-CDM predicts, she notes, and early dark energy doesn't make that problem much worse. "If you're going to say that Lyman-alpha excludes early dark energy, well, then, it also excludes Lambda-CDM," she says.

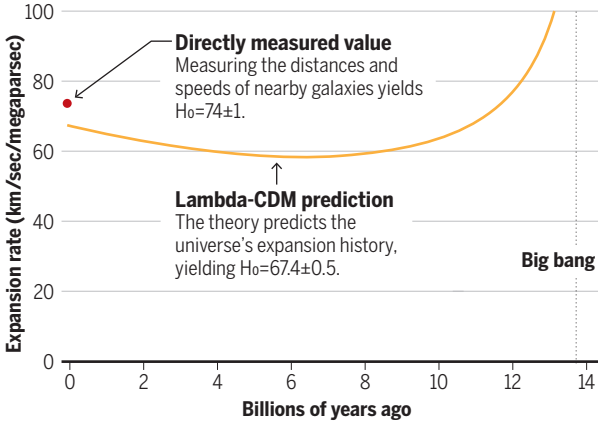
Observations that could help probe early time physics may be hard to come by. For example, the quantum field Karwal and Kamionkowski proposed might have interacted with photons and left a fingerprint in how the microwaves of the CMB are polarized, Eskilt says. He and colleagues searched for that signal in Planck's data and didn't find it, they reported on 7 September in *PRL*.

But there is a surer way to detect the field, Kamionkowski says. It should make ultrasharp pictures of the CMB differ statistically from the predictions of Lambda-CDM at angles of less than 0.2°. Future CMB studies such as the Simons Observatory, under construction in Chile, and the CMB-S4 project proposed for the South Pole should have the resolution to see such details, Kamionkowski says. "Ten years from now either we'll be seeing early dark energy or we won't be talking about it anymore."

Most cosmologists still think the Hubble tension will be resolved. "Someone will find some solution," Goldstein says. "I don't expect 100 years from now people will still be focusing on the Hubble tension." But Eskilt says the solution itself could be as mysterious as, say, dark matter. "We could be in the same position where we know that there is a source of early dark energy, but we have no idea what it is." ■

A tale of two measurements

The Hubble constant, H_0, is the rate at which the universe is expanding now. Multiplying it by the distance to a galaxy reveals how fast the galaxy is speeding away from Earth. Direct measurements of the constant yield a higher value than the one deduced from the cosmic microwave background and a theory called Lambda–cold dark matter (Lambda-CDM).



PUBLISHING

Scientists clash over 'positionality statements'

In a growing practice, researchers are adding descriptions of their identities to their papers

By **Rachel Zamzow**

When Genevieve Wojcik's coauthors suggested she include details about her race and family background in a May *Nature Genetics* commentary, she was skeptical. As a genetic epidemiologist, she had always been taught "to take yourself out of the equation completely," says Wojcik, who is at Johns Hopkins University. But Wojcik's colleagues argued that their paper, about the need for multiracial participants in genetics studies, should include a "positionality statement" from each author describing how their identity might influence their work. The practice is becoming increasingly common in scientific papers, to applause from some researchers and chagrin from others.

The statements, which can encompass anything an author deems relevant—for example, race, ethnicity, geographic location, sexual orientation, gender identity, disability status, and career level—are already established practice in many social sciences, such as sociology and anthropology. Now, they are making inroads into other fields, such as biology, global health, and medicine, as well as STEM education, with more journals encouraging or even requiring them.

"It's an invitation to think more broadly about what your role as a researcher is in the work that you're trying to understand," says Alejandra Núñez-de la Mora, a biological anthropologist at the University of Veracruz. She published a 2021 paper in the *American Journal of Human Biology* arguing that reflecting on one's positionality can pay off in future work, helping researchers address inequities such as "parachute research," unchecked power dynamics, and gaps in inclusivity. If you're an astronomer, for example, think about where your telescope is, she says. "Are you part of that community? Is that telescope put there with knowledge of the people who call that place their land?"

But others question the statements' value. "I find it amazing that [publishing positionality statements] is becoming so widespread without any evidence that it actually achieves what it sets out to achieve," says Patricia Nayna Schwerdtle, who studies global public health at Heidelberg University and coauthored a February critique of the practice in *Perspectives on Psychological Science*.

These statements—which can take a variety of formats, depending on the journal's guidelines and researchers' preferences—allow researchers to describe their identity as they see it. For Wojcik, a biracial Asian American woman, this process still felt a bit uncomfortable, as she'd previously only considered how aspects of her identity could be used against her. But despite the risk, especially for people from marginalized communities, she felt the exercise helped put her work in context.

Positionality statements also benefit readers, supporters say, peeling back the curtain on researchers' decisions that would otherwise remain invisible—from what questions they pursue to how they interpret their data. "It's not just that we want to know about people's socially constructed identities. That's not the point," says Julie Martin, editor-in-chief of the *Journal of Women and Minorities in Science and Engineering*, which has required positionality statements since July 2020. "The point is how do those and your worldview affect the decisions you make in the research?"

Alejandra Núñez-de la Mora calls positionality statements "an invitation to think more broadly."

PHOTO: ALEJANDRA NÚÑEZ-DE LA MORA

But some researchers think airing this information in the literature violates a central tenet of science: that a researcher's work should be judged independently of who they are. Spotlighting a scientist's identity represents a "bizarre turn back to [the] Dark Ages," says Anna Krylov, a chemist at the University of Southern California who wrote an April critique alongside 28 coauthors in the *Journal of Controversial Ideas*. "It was not a good time when people were treated by their attributes and not by their achievements, not by their merit," Krylov adds.

Another concern is that positionality statements serve only as virtue signaling and gloss over deeper issues, such as the reproducibility crisis in science. They seem like a "last ditch effort before you publish your paper," Nayna Schwerdtle says. Instead, researchers should strengthen upstream solutions, such as open science and participatory research, she argues.

At least one team of researchers is studying whether these statements have their intended effect. Rose Oronje, a researcher at the African Institute for Development Policy, and her colleagues at the Liverpool School of Tropical Medicine are evaluating the results of publishing reflexivity statements—a similar practice that encourages researchers engaged in global collaborations to consider how their work acknowledges the communities involved.

The team is reviewing published statements in global health journals and interviewing authors and journal editors to gauge whether this measure at the time of publication has the power to shift researchers' mindsets and lead to more equitable practices. Publication guidelines can provide powerful incentives for scientists to effect systemic change, Oronje says. "When you start there, it becomes very easy for us to want to do it, because we want to publish."

For true cultural change, proponents emphasize, these statements have to stem from ongoing self-reflection rather than a rote checklist of attributes. "I would hate for it to become a formulaic part of the writing or research process," says Núñez-de la Mora, who is editor-in-chief of the *Journal of Biosocial Science*. She and the journal's other editors are considering inviting authors to take the lead in acknowledging their identities, but not requiring it or specifying the format. The practice is catching on in her field at large, she says, especially among younger scholars.

As for Wojcik, she says reflecting on her positionality has helped her realize her identity is inextricably linked with her work, enriching it and shaping the directions it takes—including, for example, her work to design genetics studies to be more inclusive of multiracial individuals. "I've sort of come into the notion that who I am can actually make my work better and lead me to question things that hadn't been questioned before." ■

Rachel Zamzow is a freelance journalist in Waco, Texas.

CONSERVATION BIOLOGY

Plan to 'rewild' Argentina's guanacos stirs debate

Controversy includes claims of government malfeasance and threat of a defamation lawsuit

By **Kata Karáth**

A move to reintroduce guanacos—a wild relative of the llama—to a reserve in Argentina has become the latest flash point in a fierce scientific debate over the nation's rewilding efforts that has included charges of government malfeasance and legal threats.

Supporters of the effort to move 45 guanacos some 1500 kilometers north from Patagonia to a reserve in central Argentina say it could help restore the iconic, long-necked ungulate—which has seen its overall population in South America drop by 97%—to landscapes where it once thrived. But critics argue Argentina's environment ministry approved the plan without adequately evaluating its conservation value and risks, or consulting with scientists.

"This management decision violates guiding principles for rewilding," six Argentine researchers wrote last month in *Conservation Science and Practice* (*CSP*). "There is a lack of proper evaluation and scientific evidence to support the need, opportunity, and feasibility of this translocation." (The environment ministry did not respond to a request for comment.)

Earlier this year, some of the same authors were threatened with legal action after they joined dozens of other scientists in publishing a paper that critiqued Argentina's rewilding policies and a book by a staffer from ReWilding Argentina, a prominent nonprofit that has supported numerous reintroduction efforts. ReWilding Argentina's leaders complained the April critique was "written in a violent manner" and threatened to sue the authors for defamation. A guilty verdict in such a case could bar the critique's authors from working for government agencies and public universities.

This month, in a letter published in *Biological Conservation*, the scientists fired back, accusing ReWilding Argentina of trying to stifle "constructive debate. … We strongly reject this baseless threat."

ReWilding Argentina did not respond to questions about whether it plans to follow through with its legal threat. But the exchange highlighted the simmering controversy surrounding rewilding efforts in Argentina, where conservationists have moved to reintroduce jaguars, giant river otters, and other animals to former habitats. The debate intensified in September 2022 after the government approved a plan by the province of La Pampa to import 45 guanacos to its Luro reserve, which already hosts a few dozen translocated guanacos. ReWilding Argentina is providing logistical support to the effort, which some conservationists hope will open the door to translocating guanacos even farther north, to Argentina's Impenetrable National Park.

Many researchers are skeptical of the need. They note researchers estimate that more than 1 million guanacos live in Argentina, so establishing new populations "is not a conservation priority," says Ulises Balza, a biologist with Argentina's National Scientific and Technical Research Council and lead author of the *CSP* letter. And they question whether translocated guanacos, which are adapted to Patagonia's harsh environment, will thrive in their new homes, which have different ecosystems. They also fear animals imported from different populations could interbreed and reduce their genetic fitness. "The long-term viability of the population becomes uncertain," says Benito González, chair of the South American Camelid Specialist Group of the International Union for Conservation of Nature.

Critics also say the environment ministry hasn't required adequate studies before approving translocations. They note that Argentina's National Parks Administration (NPA) gave sharp scrutiny to a proposal from ReWilding Argentina to move guanacos to the Impenetrable National Park. NPA concluded more research was needed before it could advance the plan. Among the unanswered questions were whether guanacos had actually once lived in the park, and how climate change might affect the long-term viability of a new population.

In the meantime, the authors of the *Biological Conservation* letter say the flare-up has had a "silver lining" by drawing public attention to a discussion "that was previously silenced or unfolding in the shadows: the when, where, how, and why of implementing rewilding strategies." ■

More than 1.5 million guanacos live in South America. About 80% are found in Argentina.

Kata Karáth is a journalist in Quito, Ecuador.

Paraguay this year has had one of the world's largest outbreaks of chikungunya virus, which put this man in San Lorenzo in the hospital.

PHOTO: JORGE SAENZ/AP

INFECTIOUS DISEASE

A chikungunya vaccine is likely to get approved. Who will get it?

Travelers are first, but the real need is in endemic areas

By **Jon Cohen**

The first vaccine against the mosquito-borne viral disease chikungunya will likely come to market next month. With the debilitating disease now afflicting more than half the countries in the world and threatening to spread further, the imminent U.S. Food and Drug Administration (FDA) approval of the vaccine is "great news," says Scott Weaver, a virologist at the University of Texas Medical Branch whose own lab started to work on a chikungunya vaccine nearly 2 decades ago.

The vaccine, made by the French company Valneva, will likely be recommended mainly to U.S. travelers at first. But many expect an FDA approval will also grease the wheels for the vaccine to become available in the most affected countries.

First documented in 1952 in modern-day Tanzania, chikungunya means "disease that bends up the joints" in Kimakonde, an East African language. Although rarely fatal, chikungunya virus causes long-term, debilitating joint pain in up to 40% of people it infects, most of whom live in warm climates that have large populations of two mosquito species, *Aedes aegypti* and *A. albopictus*. Almost all cases in the United States have been in travelers who have returned from affected countries.

The vaccine consists of a lab-weakened version of the virus that does not cause disease, and it is likely to be approved without the usual efficacy data from human trials. FDA agreed with Valneva—and wide consensus in the field—that a standard efficacy trial comparing rates of disease in people who receive the vaccine versus dummy shots was not feasible. "The virus spreads tremendously fast," leaving too little time to launch and complete a trial, says Barbara Schnierle, a virologist at the Paul Ehrlich Institute who studies chikungunya. (Schnierle advises the European Medicines Agency about HIV vaccines but is not involved with Valneva's application there.) "Logistically, you will never be able to do a normal phase 3 study."

Valneva instead staged studies that showed 99% of people who received a single shot developed long-lasting antibodies that could neutralize the virus in test tubes. The company went on to show that when it infused monkeys with these antibodies and then gave them the virus, the animals controlled viral levels and did not develop any symptomatic disease. If the vaccine wins approval, FDA will require Valneva to conduct a postmarket study in humans to evaluate whether it actually protects against disease in the real world.

Valneva's expected approval comes after years of frustratingly slow progress. More than 5 decades ago, the U.S. Army made a chikungunya vaccine that protected monkeys and even entered a small human trial, but the military saw little need for it and didn't advance the research. Other efforts similarly languished, in part because at the time, the disease surfaced only sporadically in a few African and Asian countries.

In 2004, however, Kenya had the world's first documented large-scale outbreak of chikungunya in 30 years. The next year, a large outbreak hit Réunion Island, an overseas territory of France in the Indian Ocean that had never detected a case. Estimates suggest nearly 40% of islanders were infected, and 85% developed debilitating symptoms.

Analyses of the circulating virus showed it had mutated to allow replication in the *A. albopictus* mosquitoes, which have a different range from its original host, *A. aegypti*. The virus also became better at copying itself in both species, increasing transmission to humans. Outbreaks soon exploded in India, Thailand, the Caribbean, and Brazil. Chikungunya did so well that in many locales, it bumped out dengue virus, which the same mosquito species transmit.

To date, more than 100 countries have reported local transmission of the virus. "It just took over entire mosquito populations and caused large outbreaks," says virologist Timothy Endy, who heads a chikungunya vaccine program at the Coalition for Epidemic Preparedness Innovations (CEPI), a nonprofit that provides funding to Valneva and other developers.

Serious interest in vaccines finally ramped up—including the one made by Valneva, which it licensed from a group at the Karolinska Institute. An inactivated virus candidate from India's Bharat Biotech, which may be safer for pregnant and immunocompromised people, is close on the heels of the Valneva shot.

Last week, a chikungunya vaccines working group that advises the Centers for Disease Control and Prevention (CDC) recommended the Valneva shot primarily be offered to U.S. adults who will be visiting countries that have active outbreaks. Travelers older than 65 or with preexisting medical conditions that might make them more susceptible to severe disease also "may be considered" for a vaccine if they are visiting a country that had an outbreak within the past 5 years, the working group to CDC's Advisory Committee on Immunization Practices (ACIP) said. Also eligible: people staying in endemic countries for at least 6 months and lab workers who handle the virus as part of their research. Endy, a U.S. Army researcher for 24 years, suspects the military will recommend the vaccine for its members traveling to affected areas, because chikungunya could "bring battalions down."

Now, he says, "We need to invest in the appropriate conversations and regulatory approvals" to make the vaccine available to people beyond wealthy travelers. Among the countries that could really use the vaccine are Brazil and Paraguay, which together have seen 75% of the world's 440,000 reported chikungunya cases this year. With CEPI's help, Brazil's Butantan Institute is partnering with Valneva to bottle the vaccine and make it affordable there and to other middle- and low-income countries. At the ACIP meeting, a Valneva representative said it will likely sell for about $350 a dose to U.S. travelers. Endy expects Butantan will charge about $10 to $20 per dose.

Climate change may make chikungunya vaccines even more important. Warming climates are altering mosquito populations, which Schnierle notes are already moving north in Europe. "We've had local outbreaks in Italy, Spain, and France from travelers coming back from tropical countries," Schnierle says. "Climate change will change the whole situation." ■

EVOLUTION

Ancient fish reveal the origin of the shoulder in vertebrates

Cleft in fossil skull suggests solution to a long-standing mystery: shoulder tissue evolved from gill arches

By **Elizabeth Pennisi**

Researchers have traced an evolutionary connection between the throats of ancient fishes and the anatomy that allows you to swing a golf club or reach a high shelf. The vertebrate shoulder, they conclude, arose in part from ancestral fishes' gill arches, curved supports encircling the throat.

The analysis, reported in *Nature* this week, helps reconcile two existing ideas about where the bones connecting our arms to our trunks came from: the tissues of the head or lateral folds along ancient fishes' bodies. In the new picture, based on detailed studies of 400-million-year-old fossils, each is given a role in the evolution of paired limbs. The paper "may be a major step toward the final resolution of [this] long-standing question," says Shigeru Kuratani, an evolutionary morphologist at the RIKEN Center for Biosystems Dynamics Research who was not involved in the work.

About 500 million years ago, vertebrates were jawless marine creatures with tails but no paired fins, and numerous, paired gill arches. Fast forward 100 million years, and fish called placoderms swam the seas, with jaws and a primitive "shoulder" or pectoral girdle supporting paired front fins.

Ichthyologists have long argued that jaws arose from an ancestor's most forward gill arch. Another gill arch eventually evolved into a head bone called the hyoid. In the 1800s, biologists suggested the pectoral girdle—which eventually gave rise to the human shoulder—came from other gill arches.

Others argued the paired fins, as well as their support structures, instead arose from lateral folds, flaps of soft tissue that ran the length of a fish's body, representing trunk, not head, tissues. It makes sense that arms and legs originated in these folds, but "the girdle is more complex as it has links to head and its musculature," says Martin Brazeau, a paleontologist at Imperial College London. "How you get that is quite a mystery."

Pondering this while walking his dog, Brazeau considered CT scans of the braincase of a placoderm called *Kolymaspis sibirica*, collected more than 60 years ago in Siberia. He was intrigued by a hole, or divot, at the back of the head on each side of the braincase, which once held cartilage. He wondered whether that cartilage had once been connected to a gill arch.

Brazeau and his colleagues took a fresh look at placoderm fossils, including a close relative of *Kolymaspis* called *Brindabellaspis*. By comparing the position of the divot in *Kolymaspis* with other divots in other fossils, the researchers concluded it does represent the most tailward of a series of attachment points that once linked vertebrate braincases to their gill arches.

Known placoderms have at most five gill arches, others having evolved into jaws and the hyoid, so the pectoral girdle derives from what was once the sixth gill arch, Brazeau's team concludes. At first, the pectoral girdle enabled gill-lifting muscles to better open the mouth, leading to the evolution of diverse fish feeding systems, Brazeau says. Later it evolved to support paired fins, which he agrees came from trunk tissue.

The divot, or the area that was once the sixth gill arch, is an anatomical landmark, dividing head from trunk tissue, Brazeau and his colleagues say. Developmental studies in living fish support this idea, he says, showing that a muscle in this region called the cucullaris develops from head tissue.

"There's some tantalizing evidence in this paper," says William Bemis, an evolutionary morphologist at Cornell University. But he cautions that because "the gill arches aren't known in any of these fossils, it's all hypothetical."

Rui Diogo, a biological anthropologist at Howard University, has no doubts. "It's interesting to see a fossil so clearly showing this, as we have evidence from other fields," he says. Embryological studies by him and others show the shoulder tissue in many vertebrates, including fish and humans, develops from the same kind of tissue as the head, whereas the pelvis, arms, hands, legs, and feet all develop from trunk tissue, he says. "Independently, we have arrived at the same conclusion and this [fossil] is the most compelling direct evidence." ■

SHORE PATROL

Researchers are ringing the Atlantic with hundreds of sensors in a bid to track how ocean warming is affecting coastal ecosystems

By **April Reese**, *in Viana do Castelo, Portugal*

One wrong move, and the sea-battered granite boulders lining North Beach here can break an ankle. But this summer, marine biologist Fernando Lima hopped among the algae-covered rocks like a child playing hopscotch. He eventually settled next to a suitcase-size boulder that held what he sought: an electronic sensor, just 3 centimeters wide and shaped like a hockey puck, that he and his research team had embedded in the rock the year before.

Lima, who works with the Biopolis Association and the Research Centre in Biodiversity and Genetic Resources at the University of Porto, aimed to download data from the sensor. First he had to pry off an interloper: a limpet, a coin-size, hat-shaped snail that normally clings tenaciously to tidal zone rocks but had taken up residence on the gray epoxy seal protecting the sensor. Limpet evicted, Lima took out an iPhone and began to download data, including hourly temperature measurements recorded over the past year.

The July fieldwork was just a small part of an ambitious and unprecedented effort to deploy legions of the small sensors—which Lima's team has helped perfect—all along the shores of the Atlantic Ocean. The goal: to fill a big gap in understanding of how climate change and other factors are affecting the notoriously harsh and highly dynamic intertidal zone, where powerful waves and rising and falling tides can create a complex mosaic of microhabitats with dramatically different conditions within the same small patch of shoreline.

Satellites and airborne sensors provide only coarse measurements along these rugged continental edges, and buoys typically float too far from shore to allow accurate readings. In contrast, the on-site sensors can reveal substantial temperature differences in intertidal areas just meters apart.

Over the past 15 years, Lima and marine biologist Rui Seabra, who co-leads the Atlantic Ocean Coupled Coastal Temperature and Biodiversity Observation Network, as the project is called, have embedded nearly 1800 sensors at 162 locations along the Atlantic's eastern and western shorelines. Just this summer and fall, the researchers, with help from collaborators on both sides of the Atlantic, have installed hundreds of additional sensors along 21,500 kilometers of shoreline.

PHOTO: RUI SEABRA

Researchers check on a temperature sensor installed near fire coral on the island of Santiago in the Cape Verde archipelago.

At each site, they and their collaborators also collect images of flora and fauna, to be identified later. Ultimately, they hope to instrument 320 sites and obtain a detailed view of how thermal conditions are affecting intertidal organisms across the entire Atlantic region.

The network could provide "the seminal data set for how climate change is affecting intertidal species," says biologist Gray Williams of the University of Hong Kong. These new data could also aid efforts to protect marine biodiversity as the ocean changes, says biologist Brian Helmuth of Northeastern University, who supervised Lima when he was a Ph.D. student in the early 2000s. "These very localized measurements," he says, are "critical for any kind of management strategy."

IT TOOK YEARS for Lima and his team to perfect the loggers, the scaffolding on which the entire project hangs. The first iteration, fielded in 2008, was inspired by a sensor created by Helmuth in the late 1990s that used a real mussel shell to house some off-

Coastal sentinals

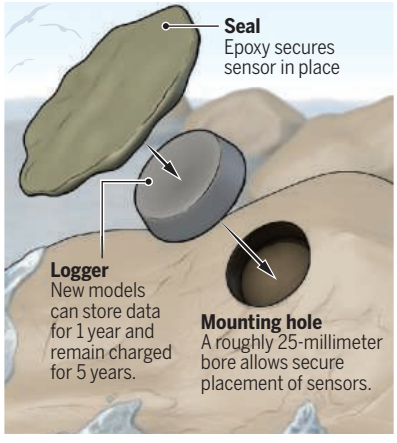
Nearly 1800 sensors (●) now monitor temperatures at about 160 locations along more than 21,000 kilometers of Atlantic coastline. The Coupled Temperature and Biodiversity Observation Network, which is still expanding, is helping scientists understand how climate change and other factors are influencing the thermal environments of marine organisms in the highly dynamic intertidal zone.

A closer look

One of the highest densities of sensors is found along the Atlantic coast of Portugal and Spain. Researchers here have been collecting data for nearly 15 years.

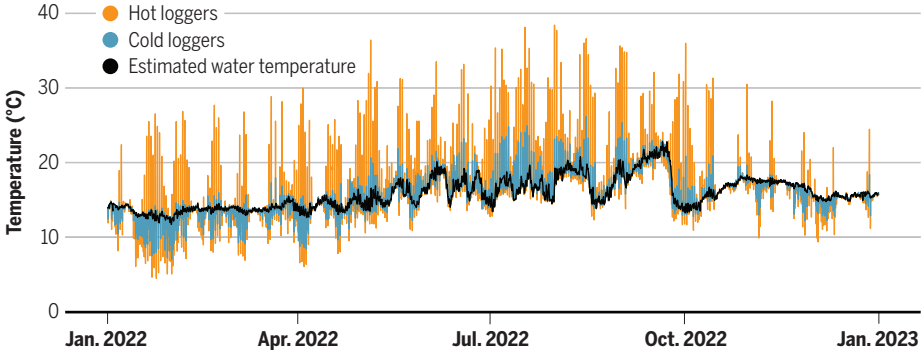
Made to last

Current sensors record temperature every 30 minutes and are designed to survive the intertidal zone's crashing surf and swirling currents.


Seal
Epoxy secures sensor in place
Logger
New models can store data for 1 year and remain charged for 5 years.
Mounting hole
A roughly 25-millimeter bore allows secure placement of sensors.

A variable environment

At a site in Portugal, the sensors revealed that intertidal temperatures vary greatly, even in locations just meters apart. Locations exposed to the Sun at low tide can be much warmer (orange bars) than less exposed nearby locations (blue bars).



the-shelf electronics. Lima opted to use a limpet shell instead, creating what he called "robolimpets."

At the time, researchers were still primarily using measurements of sea surface temperature to gauge the thermal stress experienced by organisms in the intertidal zone. But robolimpets Lima installed at 13 sites along the coast of Spain and Portugal confirmed that approach missed potentially crucial variation and underestimated how hot some animals were getting. The devices, which recorded temperatures every 30 minutes, showed that at low tide, invertebrates clinging to south-facing boulders experienced temperatures much higher than their north-facing counterparts just meters away, even at the northernmost sites. The study, published in 2011 in the *Journal of Experimental Marine Biology and Ecology*, also showed much larger daily temperature fluctuations than those inferred from sea surface readings.

Those thermal differences have implications for how different marine animals and plants might respond to warming seas, the researchers noted. Organisms living in microhabitats with higher peak temperatures, for example, might be able to adapt to warmer waters, or migrate just a short distance to survive. But those adapted to cooler temperatures might only survive if they can shift their ranges poleward.

The research also revealed some unexpected drawbacks in the early robolimpet design. At some shorelines popular with visitors, Lima recalls, "we kept losing them." The problem: "They looked like limpets," which people like to eat, he says. "So people were just collecting them all the time." Large numbers were also lost to Portugal's notoriously strong surf.

Over the next decade, Lima worked with Seabra, who first joined the project as a graduate student, to improve the sensors, enhancing data storage, battery life, and other attributes. The latest version, known as the Envlogger, sports a thick plastic housing instead of a limpet shell, can store data for a year or more, and has a battery life of 5 years. And unlike robolimpets, the Envlogger can be embedded in a rock, making it less vulnerable to would-be harvesters and waves.

As word of the Envlogger's capabilities spread, Lima and Seabra began to field requests for sensors from researchers around the world. To meet the demand and further improve the devices, in 2018 the two men hired an engineer and founded a nonprofit startup, ElectricBlue, that now sells five different loggers in various colors and shapes. The devices, which cost between €55 and €80 ($58 to $84) each, have been installed on every continent, including Antarctica.

SO FAR, FEW COASTLINES have the density of loggers found at the Viana do Castelo site, which is about an hour's drive from Lima's lab in Vairão and not far from where the 46-year-old scientist was raised. About 50 sensors, some just a half-meter apart, are spread in clusters among the rocks.

The temperature and biodiversity data they gather continue to embellish a picture of the ecosystem that Lima and his colleagues have slowly developed over nearly 15 years. During this summer's visit, Cátia Monteiro, a postdoc in Lima's lab who recently documented a drop in the abundance of cold-water algae species in this region, surveyed tangles of brown algae draped across rocks by the outgoing tide. Nearby, research assistant Mar Humet Caballero delicately pressed a tiny square black device against the shell of a limpet to "listen" to its heartbeat. The

Researcher Rui Seabra drills a hole for securely mounting a sensor near Puerto Lobos, Argentina. The devices can record hourly temperatures and store data for up to 1 year, enabling scientists to assemble a detailed picture of the intertidal zone's thermal characteristics.

instrument produces an infrared beam that passes through the shell and reflects off the snail's beating, three-chambered heart. An increased heart rate can mean the limpet is experiencing heat stress, Lima, Seabra, and colleagues noted in a 2016 study.

This and other Atlantic coast sensor sites are revealing "a whole world of thermal complexity that we were not fully aware of," Lima says. For example, the data challenge a widespread assumption, based in part on satellite data, that intertidal temperatures decrease fairly steadily as you move toward the poles. Instead, intertidal zones are a mosaic of highly variable microsites with different daily and seasonal temperature profiles.

Even at cooler, higher latitudes, some microsites "routinely experience high temperatures," Lima says. At the same time, cooler microsites are more common than once thought—even at warmer, more tropical lower latitudes. The sensors have also revealed how, during summers, the upwelling of cold water along Europe's coast—the bane of beach-goers—can help cool the intertidal zone.

"This small-scale variability has a profound impact on the way we think about why species are where they are, and what's going to happen in the future," Lima says. The cooler sites, for example, could serve as refugia for organisms that otherwise could not withstand spiking temperatures.

One of the most valuable insights the sensors could provide is which areas are most vulnerable to warming, says Williams, who has used the loggers to study heat stress in marine snails and is seeking funds to create a similar network in Asia. "I think it [can help] identify hot spots and to see where vulnerabilities will really take place," he says. These emerging hot spots are "not going to happen in the tropics—it's actually going to be further from the tropics, which is not something we would have predicted before. And this will allow us to put our focus more in areas which are really going to need conservation efforts."

WITH HELP FROM a constellation of government and philanthropic funders who collectively have contributed nearly €750,000, Lima and Seabra are now working with researchers along the Atlantic coast to fill gaps in the sensor network. The plan is for participants to upload temperature and biodiversity data into a common repository at least once a year.

Data collected by the sensors can be easily downloaded to a mobile phone.

"If we can do this in a way that's coordinated across ecosystems, we can determine what the status of biodiversity is," says Amanda Bates, a biologist at the University of Victoria who used Lima's loggers for a forthcoming study on seabirds on Newfoundland. "Ultimately what we'd really love is a weather network for nature. We need to understand what's changing."

As they expand the network, Lima and Seabra are also hoping to elevate the careers of researchers in Atlantic countries with fewer resources, particularly those in the Global South. The project's open-access data set, expected to come online in 2024, should make it easier for scientists doing localized studies to address larger questions, Lima says—and perhaps boost their chances of having their findings published.

This year, Lima and Seabra embarked on a trans-Atlantic logger blitz, covering as much coastline as possible. They replaced old sensors, installed new ones, and met with potential collaborators. Lima made his way up the eastern coast of the United States and Canada, as Seabra focused on stretches of South America, Iceland, and Scandinavia. Next stop: islands in the southern Atlantic, in March 2024. "We are focused," Lima says, "on getting this done." ■

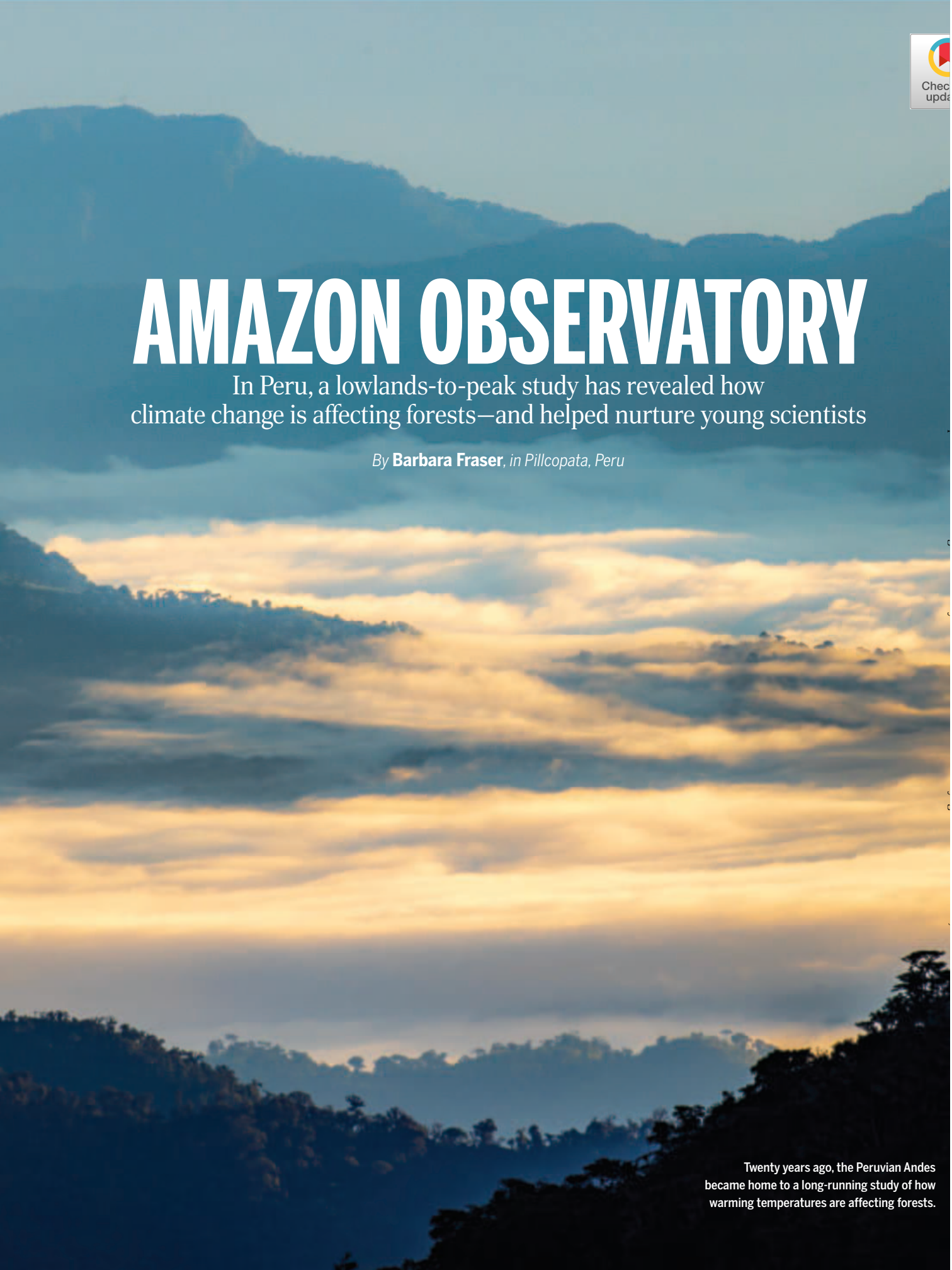
April Reese is a science journalist based in Aveiro, Portugal.

PHOTOS: (TOP TO BOTTOM) FERNANDO P LIMA; RUI SEABRA

AMAZON OBSERVATORY

In Peru, a lowlands-to-peak study has revealed how climate change is affecting forests—and helped nurture young scientists

By **Barbara Fraser**, *in Pillcopata, Peru*



Twenty years ago, the Peruvian Andes became home to a long-running study of how warming temperatures are affecting forests.

Twenty years ago, a dozen Peruvian biology undergraduates armed with machetes and tape measures laboriously cleared a trail down the steep eastern flank of the Andes Mountains near this sleepy Amazonian town. They staked out eight study plots along a 15-kilometer-long transect that stretched from the grasslands found near the relatively cool, treeless top of a 4020-meter-high mountain known as Apu Kañajhuay down the fog-shrouded Kosñipata Valley to the warmer forests below, which are soaked by as much as 5 meters of rain a year. Along the transect lies some of the world's richest biodiversity.

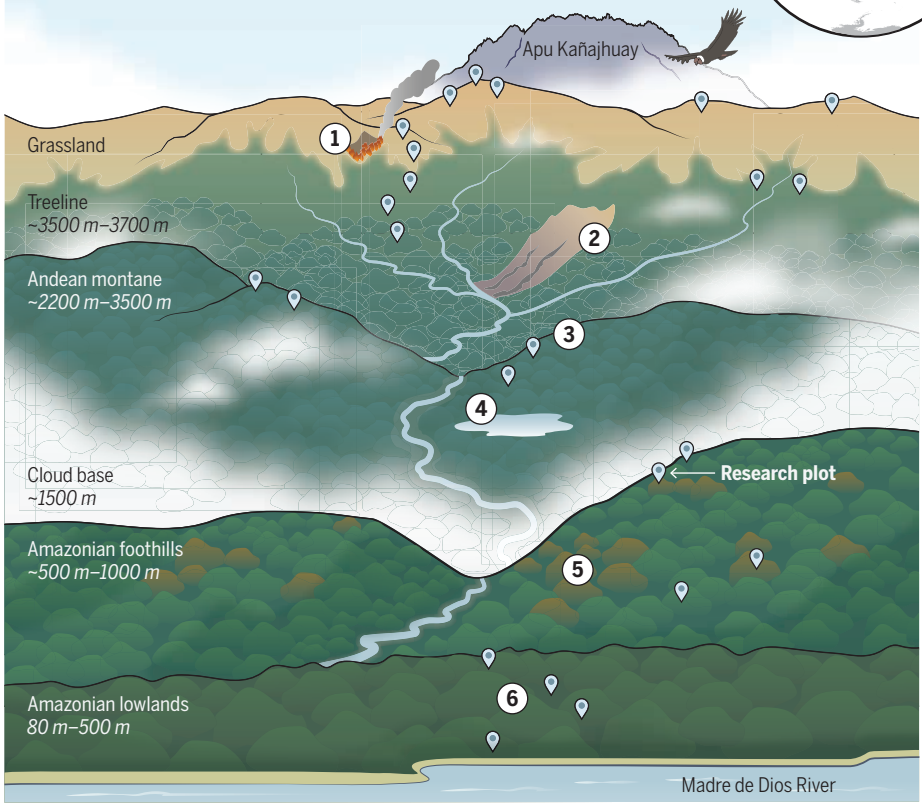
Now, those 1-hectare plots in and near Manu National Park and Biosphere Reserve are the centerpiece of one of the longest running field studies of how an Amazon forest is responding to climate change. The effort, known as the Andes Biodiversity and Ecosystem Research Group (ABERG), involves researchers from around the world. It has produced thousands of data points and scores of publications that have offered new insights into how warming and drying are reshuffling tropical ecosystems.

ABERG has also left an indelible imprint on Peru's scientific community, among the weakest in South America. Scientists from Europe and North America launched the project. But many of the students who helped hack out the original study plots went on to earn doctorates and academic posts and are now helping train a new generation of Peruvian scientists. The initiative became a "cradle of biologists," says William Farfán-Ríos, a Peruvian ecologist at Wake Forest University (WFU) who was one of those machete-wielding students.

Earlier this year, Farfán-Ríos was one of more than 30 scientists who gathered at the Manu Biological Station here to mark the 20th anniversary of ABERG. Over 3 days in June, participants celebrated the project's success at revealing an astonishingly diverse ecosystem and establishing a baseline for future studies of climate effects. But they also pondered emerging threats. The trail connecting the field plots has become a transit route for drug couriers, for example, and illegal gold mining is altering nearby ecosystems. And climate change itself may wipe out some of the biodiversity, which is still poorly understood. ABERG's "fundamental [research] questions remain the same," says Patrick Meir of the University of Edinburgh, one of the effort's original scientists. "But the context has become more urgent."

Vertical thinking

A 35-kilometer-long transect rises from hot, steamy jungle at about 200 meters to cooler grasslands near the 4020-meter peak of Apu Kañajhuay. The 24 study plots have yielded data on biodiversity and how tree species and forests along the transect are responding to warming.



1 A "grass ceiling"
Many tree species are shifting their distribution upslope at an average of 2.5 meters to 3.5 meters a year, but the tree line does not always expand into high-altitude grasslands. Grass fires, grazing livestock, frost, and difficulty in spreading seeds could be limiting upward migration.

2 Landslide response
Lidar surveys have revealed how forests recover after landslides, a major cause of tree mortality. Older landslides that keep some residual vegetation develop taller, more varied canopies.

3 Species riches
Researchers have described many new species previously unknown to science, including a tree genus, *Incadendron esseri*, that grows at 1800 meters to 2400 meters.

4 Unprecedented warming
Pollen grains and isotopes in lake sediment cores show precipitation and forest composition have fluctuated with warming and cooling cycles since the last glacial maximum 21,000 years ago. But the current rate of warming is much faster.

5 Some species can't cope
To simulate rising temperatures, researchers moved tree seedlings, including of the abundant *Weinmannia bangii*, to lower elevations. Many failed to thrive, suggesting some species might not adapt to warming.

6 Clues to the future
Studies of the functional traits of plants, such as their biochemical and morphological characteristics, could provide clues to how warming could transform the region's ecosystems.

ABERG'S ROOTS reach back to the 1990s, when a flurry of studies coming out of the Manu park revealed that the region, where the Amazon lowlands meet the Andean cloud forest, is a biodiversity hot spot. It held notably high numbers of plant species and was an ornithologist's paradise, reported researchers at the National University of Saint Anthony the Abbot in Cusco (UNSAAC) and foreign universities.

In 2002, in a bid to organize efforts to document the region, Conservation International and the Missouri Botanical Garden invited researchers from more than half a dozen universities in the United States and Europe to meet at the National Center for Ecological Analysis and Synthesis at the University of California, Santa Barbara. The scientists soon forged a loose alliance: Each group would seek funding to conduct its own studies but would share its findings—and a common research site.

Some of the researchers already had a working relationship with Norma Salinas, a botanist then at UNSAAC who was familiar

CREDITS: (GRAPHIC) N. BURGESS/*SCIENCE*; (PHOTO, OPPOSITE PAGE) MAXIME ALIAGA/NPL/MINDEN PICTURES

Ecologist William Farfán-Ríos measures a tree in the Peruvian Andes. Now at Wake Forest University, he helped establish the Andes Biodiversity and Ecosystem Research Group transect when he was an undergraduate student.

with the area targeted for study. Salinas soon became a key figure in ABERG's development. She recruited rotating brigades of students to create the study plots and spent her weekends shuttling people and supplies between Cusco and the field site, along a steep unpaved road plagued with rock falls. "She was an engine," recalls WFU tropical ecologist Miles Silman.

The project's first decade focused largely on tallying the region's enormous biodiversity. Researchers identified 1255 species of trees growing in the plots, discovered a tree genus (dubbed *Incadendron esseri*) previously unknown to science, and snapped camera-trap photos of bush dogs (*Speothos venaticus*), mountain coatis (*Nasuella olivacea*), and Andean bears (*Tremarctos ornatus*) ranging farther afield than expected.

They also traced the region's ecological history, including how ecosystems responded to warming and cooling cycles since the last ice age some 21,000 years ago. Some researchers extracted cores of sediment from lakes in the region, analyzing oxygen isotopes and pollen grains that traced how temperatures and vegetation had changed over the past 46,000 years. Those studies made it clear that modern warming in the region was occurring more rapidly than it had in the past, raising the question that has moved to the center of ABERG's work: How will climate change affect trees and other organisms along its sloping transect?

In general, climate scientists expect plant species that are adapted to warmer, lower elevations to shift upslope as higher elevations become warmer and more welcoming. But exactly how this process works—and which species will benefit or be harmed—remains murky, especially in tropical regions.

To help provide some clarity, ABERG researchers ultimately expanded their network of research plots from eight to 24. Temperatures drop by about 19°C from the lowest plot, which sits at about 200 meters, and the highest, at about 3650 meters. ABERG researchers "are really working in the most dramatic environmental gradient in the tropics," says tropical ecologist Bill Laurance of James Cook University, who is not involved in the effort. "That's the kind of place where we expect to see climate change and its broader effects manifesting, but there are tons of things about it that we don't understand yet."

AS EXPECTED, some kinds of trees are expanding their range upslope, as an ABERG team led by tropical ecologist Kenneth Feeley of the University of Miami showed in 2010. From 2003–04 to 2007–08, trees in 38 genera migrated upslope at a rate of 2.5 to 3.5 vertical meters a year, they reported in the *Journal of Biogeography*. That was a slower pace than the researchers had expected, and it raised concerns that some species would not be able to keep up with the region's warming, which has advanced by 0.5°C to 1°C over the past 20 years.

The seeds and leaves of *Incadendron esseri*, a tree genus found along the transect that is new to science.

But although trees are shifting upslope, studies show the tree line separating forest from high-altitude grassland is holding steady. A number of factors may be preventing the trees from colonizing the grasslands, researchers say, including limited frost tolerance, an inability to disperse seeds, and human activities such as grazing livestock and burning vegetation. But species unable to penetrate this "grass ceiling" may risk extinction.

A constant theme at the meeting was how much scientists still don't know. More than half of the 1255 identified tree species found in the field plots, for example, are known from just one individual. And in many cases researchers have never seen those trees flower or fruit. As result, Feeley says, "We don't know what their roles are" in the ecosystem, or how they might respond to climate change.

The researchers hope to eliminate some of those unknowns in coming years. One key question is how a warming climate could affect species' biology, and therefore their role in ecosystems. To find answers, they hope to make greater use of research technologies—such as remote sensing and genetic analyses—

The Andes Biodiversity and Ecosystem Research Group study plots sit in a biodiversity hot spot known for its rich flora and fauna, including primates.

PHOTO: ALEX HYDE/NPL/MINDEN PICTURES

that weren't widely available 2 decades ago. They would also like to replicate experiments done elsewhere that have pumped carbon dioxide into enclosed forest areas to see how plants might respond to high carbon levels in the future. So far, however, the researchers don't have the necessary funding.

ABERG researchers are also considering ways to conduct studies that could help local residents and Indigenous communities improve their lives and protect a world-renowned biodiversity hot spot. Ongoing habitat loss means "we're throwing stuff away before we know what role it plays," Feeley says, "and that's a dangerous thing to do."

ABERG'S CONTRIBUTION to strengthening Peru's scientific community has been equally important. Several of the dozen undergraduates who helped establish the early study plots were the first in their families to attend college, for example. Eight of the 12 went on to earn doctorates and now oversee research of their own. Five have returned to work in Peru, and others pursue research in Peru from universities elsewhere.

Salinas recalls that one key to recruiting those students was the paid internships she could offer thanks to ABERG funds—"something that had never been done at the university." Later, as foreign graduate students arrived, she matched them with Peruvian field assistants drawn from students she worked with. The Peruvian students soon found themselves immersed in an academic culture in which professors and graduate students welcomed their questions and ideas, unlike the formal, hierarchical style they usually encountered in their studies.

Ecosystem ecologist Walter Huaraca Huasco was one of those field assistants. He grew up in a small Andean village and spoke only Quechua until he was 7 years old. When he saw that students who spoke English had more opportunities, he learned that language the same way he had learned Spanish—by watching TV shows. Working with ABERG gave him a chance to learn to use a compass, an altimeter, and a GPS mapping unit, he recalls. "It was technology we didn't have here," says Huaraca Huasco, who went on to earn a Ph.D. at the University of Oxford, where he continues to study carbon cycling.

Richard Tito, another assistant, was the son of rural Andean farmers and spoke only Quechua until he became the first person from his community to go to college. When he was a doctoral student in Brazil, he designed an experiment based on ABERG's tree migration studies with farmers in his home community in the Apurímac region. the goal was to understand how climate change would likely affect crop plants, such as potatoes and corn. They found that a warming climate could reduce both the yield and the quality of their crops, says Tito, who is now on the faculty of the National University of San Marcos in Lima, Peru.

Tito's research, Feeley says, underscores how Peruvian scientists, particularly those who speak Quechua, can relate to local people in a way that foreign scientists cannot. Peruvian researchers also have a distinct perspective on how science can influence public policy in their home nation.

PERUVIAN STUDENTS who earn advanced degrees abroad often face difficulty returning home because of a lack of research opportunities. After she completed her Ph.D. at Oxford, Salinas found there was no job for her at her previous employer, UNSAAC. (She ultimately found one at the Pontifical Catholic University of Peru.) Such obstacles are one reason why Peru's scientific workforce has long lagged behind other South American nations. In 2020, it had just 0.44 researchers per 1000 workers, compared with the Latin American average of 2.23, notes Pedro Bernal, principal adviser to the president of CONCYTEC, Peru's national science and technology council. Still, he notes that Peru has gained ground since 2011, when that figure was 0.07.

The country is aiming higher, hoping to match the roughly nine researchers per 1000 workers found in the 38 nations that are members of the Organisation for Economic Co-operation and Development. Over the past decade, for example, the government has helped fund new positions for 172 researchers, including 102 Peruvians. Among the beneficiaries was Salinas, who in 2019 received a $1 million grant that enabled her to add six postdoctoral researchers—including five Peruvians—to her team, which was studying secondary forests in Peru.

Tatiana Boza Espinoza, who was one of Salinas's six postdocs, is now launching her own effort to strengthen Peruvian science. As a research adviser at Peru's National Institute for Research on Glaciers and Mountain Ecosystems, she is collaborating with other institutions to design a new doctoral program at the National University of San Marcos. It aims to win funding from a World Bank program that will provide grants of more than $4 million each to 20 consortia to develop graduate programs in fields related to coping with climate adaptation and information technology.

One day, some of those students might end up working with ABERG—bringing a story that began 2 decades ago full circle. In the meantime, the project's most recent class of undergraduates has just finished its work, which it presented at the June meeting. Many of the young scientists thanked the doctoral students, who often had once been field assistants themselves, for mentoring and advice. It was a reminder of how ABERG has helped bolster Peruvian science, Feeley says. "I love to see the students who are finishing their Ph.D.s—who started in the field, studying English with a headlamp—[who] are now professors and publishing," he says. "The bar has been raised." ■

Barbara Fraser is a freelance journalist in Peru.

A boy wades in what is left from a kilometer-wide affluent of the Rio Negro River, near Manaus, Brazil, as the region is hit by a severe drought.

ENVIRONMENTAL SCIENCE

Water scarcity is exacerbated in the south

The Southern Hemisphere has experienced a 20% drop in water availability in 20 years

By **Günter Blöschl**[1] *and* **Pedro L. B. Chaffe**[2]

Water availability—the difference between precipitation and evaporation—is expected to decrease in many regions of the world because climate change has modified the water cycle, more water will be abstracted by the growing population, and water pollution will continue to limit clean water (*1*). Estimates of changes in water availability are uncertain at continental and global scales because measurements of precipitation and evaporation tend to be indirect or only locally representative (*2*). On page 579 of this issue, Zhang *et al.* (*3*) report a study that combines streamflow observations of large river basins of the world with terrestrial precipitation data and satellite measurements of evaporation and water storage to show that water availability in the Southern Hemisphere has substantially decreased from 2001 to 2020. With their approach, the authors have improved the reliability of water availability estimates, which could help improve long-term water management.

Zhang *et al.* used a consistent method to validate ensembles of water availability estimates. For the Northern Hemisphere, the authors find no change in average water availability from 2001 to 2020, whereas for the Southern Hemisphere, water availability decreased by 70 mm per year, which corresponds to a reduction of about 20% over the entire period. This means that the global trend in water availability is brought about by changes in the Southern Hemisphere. Similarly, the year-to-year variability in water availability is mainly caused by changes in the south. According to the authors, in arid regions of the Southern Hemisphere, these changes are predominantly related to increased evaporation, whereas in humid regions in the south, they are mainly related to decreased precipitation. Both trends and the year-to-year variability of water availability are aligned with climate modes (long-term variations of the climate that are related to ocean processes), such as El Niño–Southern Oscillation (ENSO). ENSO can trigger droughts and floods in certain parts of the world. For example, in 2023, droughts hit Amazonia while, at the same time, Southern Brazil suffered floods.

The findings of Zhang *et al.* for the past two decades suggest future water management challenges in the Southern Hemisphere. Long-term water availability, as reflected in streamflow and groundwater recharge, is the base value against which to

[1]Institute of Hydraulic Engineering and Water Resources Management, Technische Universität Wien, Vienna, Austria. [2]Department of Sanitary and Environmental Engineering, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil. Email: bloeschl@hydro.tuwien.ac.at

assess planned water abstractions for irrigation, industry, and households while leaving water available for ecosystems. When water availability in streams and groundwater drops below water demand, drought conditions are felt by ecosystems and people. This was the case during 1996 to 2010 in Australia (*4*), 2013 to 2015 in South America (*5*), 2015 to 2018 in South Africa (*6*), and during several other droughts in the past two decades in the Southern Hemisphere.

The consequences of declining water availability over decadal scales are manifested in decreasing streamflow and groundwater levels in vast swaths of land. For example, substantially decreasing trends of streamflow have been observed in large parts of South America (*7*). Water availability fluctuations must also be considered at shorter timescales (months). In regions with seasonal rainfall, evaporation may quickly dry out the soil at the onset of the dry season, leading to flash droughts (*4*). Counterintuitively, in a drier climate, rainfall may be more concentrated in the wet season (*2*), leading to floods instead of groundwater replenishment. More droughts and floods represent an acceleration of the terrestrial part of the water cycle (faster storage and movement of water between land, ocean, and atmosphere), leading to increased ecosystem degradation through tree mortality and thus greater carbon dioxide emissions. This situation has been taking place in Amazonia (*8*), further intensifying climate change effects.

The impact of droughts and floods on humans have been enormous, affecting more than 3 billion people with damages exceeding $780 billion US dollars worldwide in the past two decades (*1*). Measures to mitigate their effects include construction of infrastructure, such as storage dams and diversions for irrigation; nature-based solutions, including enhancing infiltration of rainwater into the soil; and soft measures, such as water savings through crop choice and raising awareness. However, human response to water stress may have unexpected repercussions on the water cycle. In parts of South America, increased water use for agriculture has contributed 30% to rising drought trends of streamflow (*7*). More generally, water stress in semiarid regions globally may be amplified by greater irrigation demand under climate change (*2*).

Long-term feedbacks between humans and water in catchment areas may further exacerbate water shortages (*9*). Water reservoir construction tends to increase water demand, leading to the paradoxical effect that society becomes more vulnerable rather than more resilient. Examples of this are the 2015 to 2018 Cape Town drought (*6*) or even the water scarcity that hit the ancient Maya in the 9th century (*10*). In both cases, water consumption was far above natural water availability without reservoirs. Similarly, the construction of levees—embankments built along river edges—tends to increase vulnerability to large floods because citizens are enticed to move into flood plains (*11*). Comparative analyses of droughts and floods in different parts of the world have shown that societies reduce their vulnerability only if events of similar magnitudes have been experienced in the past (*12*). These long-term feedbacks are of particular interest in vulnerable parts of the Southern Hemisphere, where many countries are mid- to low-income, with limited technological and financial means that would help them overcome environmental crises.

There are water management challenges ahead brought about by reduced water resources in the Southern Hemisphere. Reduced water availability requires a shift from crisis response to long-term, proactive water management, as advocated in the Prague Statement of the International Association of Hydrological Sciences (IAHS) (*13*). This proactive water management needs to be aligned with global goals, such as the United Nation's Sustainable Development Goals, while also drawing on the expertise of local citizens, hydrologists, and water managers. Strategies may rely on robust solutions under different possible futures that include long-term feedbacks of the coupled human-water system. These solutions may include diversifying water supply and flood protection systems and planning for flexibility in water use to reduce the potential impact of extreme events. ■

REFERENCES AND NOTES

1. UNESCO World Water Assessment Programme, *The United Nations World Water Development Report 2023: Partnerships and Cooperation for Water* (United Nations, 2023).
2. B. F. Zaitchik, M. Rodell, M. Biasutti, S. I. Seneviratne, *Nat. Water* **1**, 502 (2023).
3. Y. Zhang *et al.*, *Science* **382**, 579 (2023).
4. X. Yuan *et al.*, *Science* **380**, 187 (2023).
5. H. Escobar, *Science* **347**, 812 (2015).
6. M. Muller, *Nature* **559**, 174 (2018).
7. V. B. P. Chagas, P. L. B. Chaffe, G. Blöschl, *Nat. Commun.* **13**, 5136 (2022).
8. D. M. Lapola *et al.*, *Science* **379**, eabp8622 (2023).
9. M. Sivapalan, H. H. G. Savenije, G. Blöschl, *Hydrol. Processes* **26**, 1270 (2012).
10. L. Kuil *et al.*, *Ecol. Econ.* **157**, 1 (2019).
11. G. F. White, "Human adjustment to floods," thesis, University of Chicago, Chicago, IL (1945).
12. H. Kreibich *et al.*, *Nature* **608**, 80 (2022).
13. S. Ceola *et al.*, *Hydrol. Sci. J.* **61**, 2803 (2016).

ACKNOWLEDGMENTS

G.B. is supported by European Research Council (ERC) Advanced Grant "FloodChange," project no. 291152. P.L.B.C. is supported by the National Council for Scientific and Technological Development (CNPq).

10.1126/science.adk8164

BATTERIES

Solid electrolytes redefine ion conduction

A mechanism of ion transport in solid electrolytes can guide the design of lithium batteries

By **Anton Van der Ven**

Common lithium (Li)–ion batteries have profoundly transformed modern society by altering how energy is generated and consumed. Every charge and discharge cycle of a battery requires shuttling Li ions over large distances—micrometers to millimeters. Commercial Li-ion batteries contain a liquid electrolyte to facilitate the rapid transfer of Li ions between the anode and the cathode, but there is a strong incentive to replace this with a solid electrolyte. Solid-state batteries promise to be safer and more compact than those with liquid electrolytes and could enable a substantial increase in the energy density of Li-ion batteries (*1*). However, ion transport through most solids is sluggish compared with that in liquids. On page 573 of this issue, Yu *et al.* (*2*) provide a detailed study on Li-ion transport optimization in the solid state, which will guide the development of a promising class of solid–electrolyte batteries.

The fastest solid-ion conductors are metal sulfides and oxides, but those compounds that exhibit high Li-ion conductivities are susceptible to decomposition reactions when placed in physical contact with the cathode or anode (*1*). Moreover, many metal oxides and sulfides do not possess suitable mechanical properties to accommodate the cyclic stresses that solid-state batteries undergo as a result of volumetric changes of the cathode during cycling (*1*). This has spurred interest in solid electrolytes such as Li-metal-halides, e.g., Li_3MX_6 (where M is a metal and X is a halogen). These compounds adopt crystal structures that can conduct Li ions. Li-metal-halides are electrochemically more stable than their metal sulfide counterparts and are more

ductile than metal oxide fast-ion conductors (*3*). Li-metal-halides have been known for several decades, but it is only recently that Li-ion conductivity has been optimized to approach those of the best metal sulfide superionic conductors (*4*).

The rational design of fast-ion–conducting materials, such as Li-metal-halides, requires precise understanding of ion-transport mechanisms. The rich diversity of crystal structures among candidate electrode and electrolyte materials has revealed an equally rich variety of ion-transport mechanisms (*5*–*8*). Ion transport in most crystals is mediated by exchanges (or hops) of ions to empty (cation-free) sites, called vacancies. This mechanism is usually slow because it requires thermal activation to overcome the energy barriers to each hop. The fastest ion-conducting solids, by contrast, tend to exhibit liquid-like transport of ions along interconnected tunnels within a crystal structure. Much progress has been made to understand the chemical and crystallographic factors that foster liquid-like, fast-ion conduction, and consequently, powerful design principles have been formulated to guide the search for new fast-ion–conducting solids (*6*, *9*–*11*). None of the proven mechanisms of fast-ion conduction, however, suggests that Li-metal-halides would be a promising solid electrolyte. Yet, recent studies have demonstrated that Li-metal-halides can achieve Li-ion conductivities of 1 mS/cm (*4*). Conductivities exceeding 10 mS/cm have been achieved in some solid-state electrolytes, albeit sacrificing stability and precluding widespread use (*1*).

Most solid electrolytes are made of intertwined lattices of different chemical species. Negatively charged anions (oxygen, sulfur, or halide ions) tend to be the larger species and often arrange themselves in a crystal that can host positively charged cations (metals) within interstitial sites, the void pockets within an anion crystal. When the anions form a structure similar to that of close-packed spheres, as in Li-metal-halides, the cations fill interstitial sites that are either tetrahedrally or octahedrally coordinated by anions. Li-ion transport in these structures usually occurs by a succession of hops between octahedral sites that pass through a tetrahedral site (*5*) (see the figure). An octahedral-to-tetrahedral-to-octahedral hop mechanism favors clusters of vacant sites as the prime mediator of ion transport (*5*). The mechanism crucially depends on the availability of fully "connected" networks of tetrahedral sites that are each surrounded by multiple vacancies, forming percolating pathways of low–energy barrier hop environments (*12*). Ion transport by this means can become very sluggish when too many other cations are present (*13*).

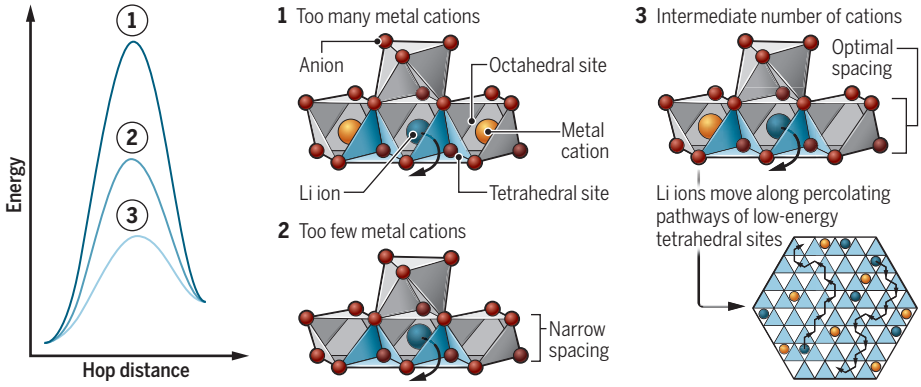
The study of Yu *et al.* establishes a fundamental understanding of the ion-transport mechanisms that enable high conductivities in Li-metal-halides. By combining molecular dynamics simulations with a careful analysis of energy barriers and their dependence on lattice shape and size, the authors show that the bottleneck of Li-ion transport in Li-metal-halides is the tetrahedral sites located between neighboring octahedral sites. Yu *et al.* demonstrate that the energies of tetrahedral sites are very sensitive to the presence of neighboring cations and the spacing between close-packed anion layers. Both are affected by the number of cations in the compound, but in opposing ways: Fewer cations lead to percolating pathways of favorable tetrahedral sites, whereas more cations serve as pillars to maintain an optimal interlayer spacing (*14*).

The findings of Yu *et al.* show that reducing the number of Li and metal cations through aliovalent doping (the introduction of ion impurities with different oxidation states) of Li_3MX_6 (e.g., replacing a fraction of $M = Y^{3+}$ with Zr^{4+}) opens up percolating networks of vacancy clusters to facilitate low–energy barrier hops. However, owing to the sensitivity of the tetrahedral site energies to the lattice geometry, there is a limit to the number of cation vacancies that should be created through aliovalent doping. A sufficient number of cations are necessary to provide enough interlayer space between close-packed anion layers to ensure a low energy barrier for hops through tetrahedral sites. These competing design principles lead to an optimal cation-to-anion ratio that Yu *et al.* used to synthesize a proof-of-concept solid electrolyte with an exceptional Li-ion conductivity.

Although high conductivities on the order of 1 mS/cm have been achieved in similar Li-metal-halides (*4*), and past modeling studies have shown the importance of reducing the cation content to increase ion conductivities in these materials (*15*), the findings of Yu *et al.* enable the articulation of precise design principles to further improve these promising materials (*3*). A clear understanding of ion-transport properties in Li-metal-halides can help achieve widespread commercialization of all solid-state batteries. ■

Lithium diffusion through a solid electrolyte

Lithium (Li) ions hop through a solid Li-metal-halide electrolyte, which forms an anion lattice with metal cations located within octahedral sites. With too many cations, the hopping energy barrier is high and Li transport is slow (**1**). With too few cations, the spacing between anion layers shrinks, impeding transport (**2**). With an intermediate amount of cations, percolating pathways of low–energy barrier tetrahedral sites open up, leading to higher Li mobility (**3**).



Materials Department, University of California Santa Barbara, CA, USA. Email: avdv@ucsb.edu

REFERENCES AND NOTES

1. T. Famprikis, P. Canepa, J. A. Dawson, M. S. Islam, C. Masquelier, *Nat. Mater.* **18**, 1278 (2019).
2. S. Yu *et al.*, *Science* **382**, 573 (2023).
3. J. Liang, X. Li, K. R. Adair, X. Sun, *Acc. Chem. Res.* **54**, 1023 (2021).
4. K.-H. Park *et al.*, *ACS Energy Lett.* **5**, 533 (2020).
5. A. Van der Ven, J. Bhattacharya, A. A. Belak, *Acc. Chem. Res.* **46**, 1216 (2013).
6. X. He, Y. Zhu, Y. Mo, *Nat. Commun.* **8**, 15893 (2017).
7. Z. Zhang, L. F. Nazar, *Nat. Rev. Mater.* **7**, 389 (2022).
8. J. L. Kaufman, A. Van der Ven, *Chem. Mater.* **34**, 1889 (2022).
9. Y. Wang *et al.*, *Nat. Mater.* **14**, 1026 (2015).
10. D. Di Stefano *et al.*, *Chem* **5**, 2450 (2019).
11. K. Jun *et al.*, *Nat. Mater.* **21**, 924 (2022).
12. J. Lee *et al.*, *Science* **343**, 519 (2014).
13. S. K. Kolli, A. Van der Ven, *Chem. Mater.* **33**, 6421 (2021).
14. K. Kang, Y. S. Meng, J. Bréger, C. P. Grey, G. Ceder, *Science* **311**, 977 (2006).
15. Y. Liu, S. Wang, A. M. Nolan, C. Ling, Y. Mo, *Adv. Energy Mater.* **10**, 2002356 (2020).

10.1126/science.adk8617

DEVELOPMENT

Sex-biased gene expression in mammals

Sex differences in gene expression start at puberty and vary across species and organs

By **Marie Sémon**

Among many mammals, males and females of the same species look different—a phenomenon called sexual dimorphism. This includes differences in morphology, physiology, and behavior. Sexual dimorphism has long been a topic of interest for evolutionary biologists and ecologists (*1*). Around 10 years ago, it was acknowledged that disease and drug studies should account for sex, making this topic a focus in biomedical research too (*2*). Sexual dimorphism is hypothesized to originate in early embryos, but data from prenatal development were lacking. On page 533 of this issue, Rodríguez-Montes *et al.* (*3*) quantify sex bias in gene expression throughout development in six species—five mammals and a bird. They found that sex differences arise around sexual maturity, thus refuting the early-onset hypothesis. They also report that sex-biased gene expression has rapidly evolved, with consequences for understanding evolutionary mechanisms and for the use of preclinical data to make predictions about sex-specific effects of treatments in humans.

In many species, male and female embryos exhibit different susceptibilities to environmental factors during early gestation (*4*), and many diseases with origins during embryogenesis exhibit sex biases in mice (*2*). Furthermore, many mouse lines with experimentally induced gene deletions show clear differences between the sexes after controlling for body size (*5*). Therefore, sexual differentiation was hypothesized to be a progressive developmental process.

Sexual dimorphism can be quantified by counting sex-biased genes—either genes with quantitatively different expression in males versus females or, more rarely in mammals, genes expressed exclusively in one sex (*1*, *6*). Rodríguez-Montes *et al.* aimed to quantify and date the onset of sexual dimorphism in mammals. They collected a large gene expression dataset from the brain, cerebellum, heart, kidney, and liver of both sexes in humans (*Homo sapiens*), mice (*Mus musculus*), rats (*Rattus norvegicus*), rabbits (*Oryctolagus cuniculus*), opossums (*Monodelphis domestica*), and chickens (red junglefowl, *Gallus gallus*) at 5 to 16 time points from early organogenesis to adulthood. They then identified the onset of sex bias for the expression of each gene tested. There were relatively low amounts of sex-biased expression in embryogenesis and a sharp increase around sexual maturity.

Two scenarios could explain this observation. Either the sexual identity of organs is achieved by sexual hormones at puberty, or it is slowly built from embryogenesis onward by small changes in the expression of many genes, which were undetected by Rodríguez-Montes *et al.* Gene expression was measured from whole organs, so it is also possible that patterns of sex-biased gene expression were masked by heterogeneity in cell type abundance between the sexes, as was recently shown in the guppy (*Poecilia reticulata*) (*7*). The organs studied by the authors do not carry a priori major sexual characteristics, so studying gonads or direct targets of sexual selection may result in different findings.

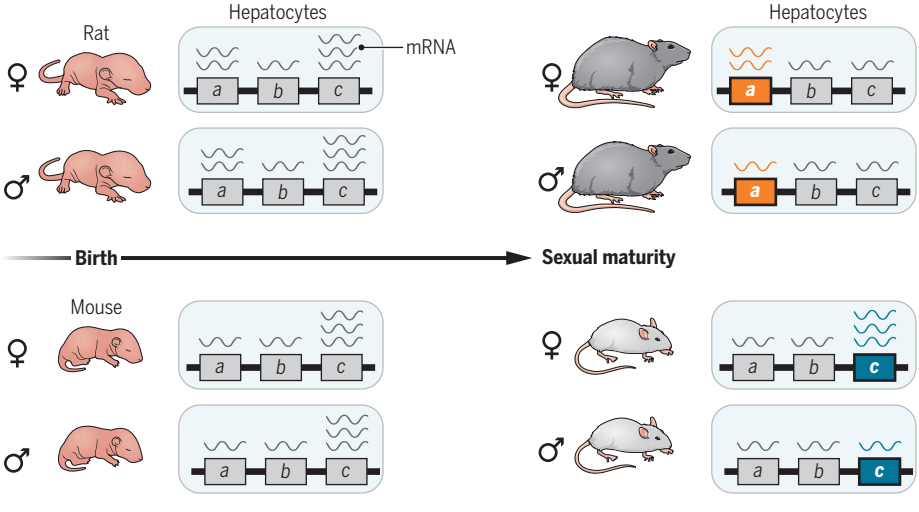
Rodríguez-Montes *et al.* found that sexual dimorphism did not have the same genetic basis in the different species studied. This indicates quick evolution of these traits, as observed previously between humans, macaques (*Macaca fascicularis*), mice, rats, and dogs (*Canis familiaris*) (*8*) and at a shorter evolutionary scale in the experimental evolution of sexual selection in *Drosophila pseudoobscura* (*9*); however, the underlying mechanisms are still unclear. Rapid evolution may be due to random genetic drift or to positive selection triggered by a conflict between the evolutionary interests of the two sexes.

Rodríguez-Montes *et al.* found that in the liver and kidney, the same cell types (hepatocytes and proximal tubule cells, respectively) were enriched for sex-biased gene expression in different species but the specific genes showing sex-biased expression differed across species (see the figure). Performing functional tests of sex-biased genes in closely related species, such as mice and rats, would allow the assessment of whether the species-specific gene expression changes correspond to the evolution of sexual dimorphism in terms of morphology, physiology, or behavior.

Historically, only male animals were used in preclinical research, and male participants were enrolled in clinical trials

Development of sex-biased gene expression

During prenatal development and at birth, there are low levels of sex-biased gene expression in organs such as the liver in mammals, including mice and rats. Around sexual maturity, there is a sharp increase in sex-biased gene expression. The same cell types are affected by this change in different species (hepatocytes in the case of the liver), but the specific genes with sex-biased expression differ.



GRAPHIC: K. HOLOSKI/*SCIENCE*

Laboratory of Biology and Modeling of the Cell, École Normale Supérieure de Lyon, CNRS UMR5239, Inserm U1293, University Claude Bernard Lyon 1, Lyon, France. Email: marie.semon@ens-lyon.fr

(*10*). These choices were justified by the assumptions that the results would be equally applicable to females or that accounting for different sizes in males and females would be sufficient to adapt the treatments to females (*2*). However, the prevalence and severity of many common human diseases differ among sexes, and many drugs have sex-specific effects (*11*).

Large-scale transcriptome studies, such as that of Rodríguez-Montes *et al*., are useful because they can guide the interpretation of preclinical and clinical trial data by providing repositories of sex-biased genes in different organs, both in the model species used in preclinical studies and in humans. However, most sex-biased genes were species specific, even in closely related species such as the mouse and rat, and the trajectories of expression during development of thousands of genes (including disease-related genes) differ between human and other mammalian species (*12*). Therefore, it will be difficult to apply knowledge on sex biases found in preclinical animal studies to predict sex biases in clinical trials.

So far, large datasets quantifying gene expression throughout development in several species have been used in the study of evolution to quantify conservation and divergence at different time points (*13*–*15*). Rodríguez-Montes *et al*. also integrated data on epigenetic modifications, single-cell transcriptomes, and genes targeted by hormone receptors to provide a functional interpretation of their findings. To understand how sexual dimorphism develops and evolves, this integrated approach should be used to study species and organs for which there is good knowledge of life history traits and morphology under sexual selection (*1*). ■

REFERENCES AND NOTES

1. N. M. Tosto, E. R. Beasley, B. B. M. Wong, J. E. Mank, S. P. Flanagan, *Nat. Ecol. Evol.* **7**, 981 (2023).
2. N. A. Karp *et al*., *Nat. Commun.* **8**, 15475 (2017).
3. L. Rodríguez-Montes *et al*., *Science* **382**, eadf1046 (2023).
4. P. J. Hansen, K. B. Dobbs, A. C. Denicol, L. G. B. Siqueira, *Cell Tissue Res.* **363**, 237 (2016).
5. W. van der Bijl, J. E. Mank, *Evol. Lett.* **5**, 359 (2021).
6. M. Oliva *et al*., *Science* **369**, eaba3066 (2020).
7. I. Darolti, J. E. Mank, *Evol. Lett.* **7**, 148 (2023).
8. S. Naqvi *et al*., *Science* **365**, eaaw7317 (2019).
9. P. Veltsos *et al*., *Mol. Ecol.* **31**, 3374 (2022).
10. I. Zucker, A. K. Beery, *Nature* **465**, 690 (2010).
11. S. L. Klein, K. L. Flanagan, *Nat. Rev. Immunol.* **16**, 626 (2016).
12. M. Cardoso-Moreira *et al*., *Cell Rep.* **33**, 108308 (2020).
13. M. Cardoso-Moreira *et al*., *Nature* **571**, 505 (2019).
14. H. Zalts, I. Yanai, *Nat. Ecol. Evol.* **1**, 0113 (2017).
15. A. D. Cutter, R. H. Garrett, S. Mark, W. Wang, L. Sun, *Evol. Lett.* **3**, 359 (2019).

ACKNOWLEDGMENTS

Thanks to S. Pantalacci for all the discussions.

10.1126/science.adk9990

ECOLOGY

Trophic cascades help restore vegetation

Herbivores and their predators have a major impact on restoration outcomes

By **Nacho Villar**

Herbivores engineer ecosystems through top-down control of vegetation, as well as by regulation of plant diversity and dispersal, fire severity, soil nutrient cycles, and carbon stocks, among other processes (*1*, *2*). For years, the impact and potential role of reintroducing, facilitating, or preventing herbivores from accessing areas where restoration of lost or degraded vegetation is encouraged has been debated. Resolving this question is central to achieving the ambitious targets of the United Nations Decade on Ecosystem Restoration initiative, which recognizes the critical role of wildlife in preventing the decline of biodiversity, restoring food webs, and promoting climate change mitigation (*3*). On page 589 of this issue, Xu *et al*. (*4*) present evidence that herbivores have a strong impact on vegetation restoration and provide insights that could accelerate restoration efforts.

In their natural habitats, herbivores reduce the abundance and biomass of plants. In terrestrial environments, this usually increases plant diversity. Xu *et al*. analyzed evidence from thousands of studies across tropical, subtropical, and temperate terrestrial and aquatic ecosystems, and their results support this view for relatively undisturbed terrestrial ecosystems. In undisturbed aquatic ecosystems, they found a larger net impact of herbivory on vegetation biomass and a null net effect on plant diversity. These results align with previous studies that have shown that herbivory can increase or decrease plant diversity in aquatic ecosystems, depending on the context (*5*).

To understand the effects of herbivores on vegetation restoration, Xu *et al*. compared the effects of herbivores on relatively undisturbed terrestrial and aquatic sites with their effects on sites where restoration of lost vegetation was being encouraged. In contrast to undisturbed sites, herbivores on restoration sites had a net negative effect on diversity and a more negative effect on the abundance of plants. Furthermore, the authors showed that such negative effects are, in general, larger for sites where restoration is actively promoted through plantation of vegetation (usually a more expensive and logistically costly option) than for sites where restoration is passively achieved through natural regeneration. Herbivory also slowed down plant recovery, had a more negative effect on the abundance of native plant species than on exotic species, and, in general, had more acute effects on tropical and subtropical zones than on temperate ones. These results indicate that managing herbivory might increase the effectiveness of vegetation restoration globally. In terrestrial systems, this conclusion suggests that the carbon sequestration initiatives that are being run in tropical and subtropical forests might also benefit from managing herbivory to boost aboveground carbon stocks.

Yet Xu *et al*. also show that herbivore biomass per square meter in natural regeneration restoration areas is approximately twice that of undisturbed sites, and, for planted restoration sites, it is three times higher. These excesses in herbivore biomass also came with shifts in herbivore composition. Compared with undisturbed sites, sites undergoing passive natural regeneration appeared to hold a substantial excess of large and especially small vertebrate biomass, and planted restoration areas had a threefold excess of invertebrate biomass.

Collectively, these findings suggest that the undesired outcome of herbivory in restoration interventions is directly linked to compositional imbalances in herbivore communities and concomitant excesses in herbivore biomass per unit area. This could be due to high densities of semidomestic livestock in some sites (a common practice in passive restoration interventions), a lack of carnivore regulation of wild large herbivore densities, or, in other sites, the proliferation of small mammals and invertebrates in the absence of large herbivores (*1*). Regardless of the cause, the findings indicate that vegetation restoration interventions often fail to restore the trophic structure, food webs, faunal compositions, and top-down regulation of the ecosystems that they try to replicate.

Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands. Email: n.villar@nioo.knaw.nl

This conclusion challenges the common perception that vegetation restoration often results in faunal restoration through passive re-colonization (*6*) and highlights that ecosystem disassembly and reassembly are not symmetrical. In terrestrial systems, for example, many wild animals are not likely to recolonize (relatively small) restored vegetation patches that are embedded in highly fragmented landscapes. This could be due to barriers to movement (*7*) or to local, regional, or functional extinctions (*1*, *8*).

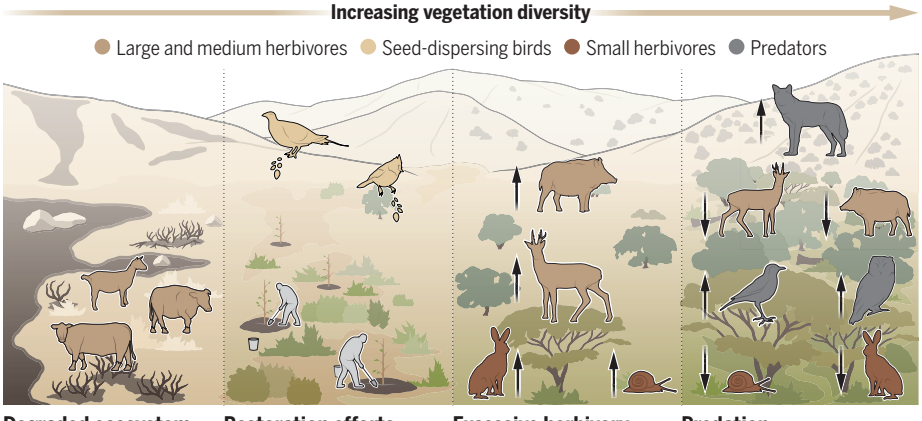
An established principle in ecology is that, through trophic and behavioral cascades, predators limit herbivore densities, which reduces consumption of vegetation and thus increases vegetation biomass (*9*). Xu *et al.* reviewed the few available studies of the impact of predator reintroductions on vegetation restoration. The results suggest that predator reintroduction might be as effective for vegetation restoration as removing herbivores (through physical exclosures, insecticides, or deterrents) or plant-centered solutions (such as decreasing plant competition or promoting facilitation), indicating that the presence of predators in an ecosystem plays a critical role in the restoration of its vegetation (see the figure). A major benefit of this nature-based solution is that it enables vegetation restoration while also restoring the trophic structure and potentially the composition and food webs of degraded ecosystems.

All predator reintroduction studies analyzed by Xu *et al.* come from temperate terrestrial ecosystems, where wildlife, in general, and predators, in particular, are slowly returning (*10*). The findings suggest that this could accelerate vegetation regeneration. In tropical and subtropical systems, predator and herbivore species whose numbers are declining (*9*) could be reintroduced for concomitant restoration of vegetation and trophic structure. Yet empirical field studies need to be conducted to validate this hypothesis.

Although the study of Xu *et al.* supports the importance of trophic cascades and the key role of trophic rewilding in ecosystem restoration, the specific restoration solutions applied should also consider local contingencies and the broader socioeconomic context. Their work also emphasizes that restoration of the diversity in Earth's ecosystems is intrinsically linked to the restoration of their biostructure. ■

Restoring food webs to regenerate vegetation

Restoration of ecosystems such as forests can be initiated by the active planting of vegetation or by natural processes such as seed dispersal by birds. However, excessive numbers of herbivores and imbalances in the composition of herbivore communities can limit regeneration through high levels of herbivory. Reintroducing predators to the ecosystem can control herbivory and facilitate restoration.



Degraded ecosystem
Vegetation is lost through fire, deforestation, and livestock overgrazing.

Restoration efforts
Planting of diverse vegetation and natural regeneration are steps toward restoration.

Excessive herbivory
Excessive numbers of herbivores consume grass and young trees.

Predation
Reintroduction of predators controls herbivory and increases trophic diversity.

GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. R. Dirzo *et al.*, *Science* **345**, 401 (2014).
2. O. Schmitz *et al.*, *Nat. Clim. Change* **13**, 324 (2023).
3. United Nations (UN), "UN Decade on Ecosystem Restoration: Action plan booklet" (UN, 2023).
4. C. Xu *et al.*, *Science* **382**, 589 (2023).
5. E. S. Bakker *et al.*, *Aquat. Bot.* **135**, 18 (2016).
6. L. Poorter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2003405118 (2021).
7. M. A. Tucker *et al.*, *Science* **359**, 466 (2018).
8. A. Benítez-López *et al.*, *Science* **356**, 180 (2017).
9. M. L. Pace, J. J. Cole, S. R. Carpenter, J. F. Kitchell, *Trends Ecol. Evol.* **14**, 483 (1999).
10. S. E. H. Ledger *et al.*, "Wildlife comeback in Europe: Opportunities and challenges for species recovery" (Rewilding Europe, 2022).

10.1126/science.adl0578

NEUROSCIENCE

The neural basis of mental navigation in rats

A brain–machine interface demonstrates volitional control of hippocampal activity

By **Michael E. Coulter**[1] *and* **Caleb Kemere**[2]

Complex behaviors such as navigation often require chains of decisions that can be mentally simulated ahead of carrying out the behavior. The activity of spatially coding neurons (place cells) in the hippocampus and associated cortical regions has been posited as the expression of a "cognitive map," which could serve as a substrate for generating these complex behaviors (*1*). On page 566 of this issue, Lai *et al.* (*2*) report that rats can navigate through a virtual reality (VR) environment using a brain–machine interface (BMI) that estimates the rat's mental location from the ongoing activity of neurons in its hippocampus. This finding is an exciting expansion of BMI implementation from sensorimotor functions to a more cognitive domain and suggests that hippocampal activity is under volitional control. In addition, the BMI approach provides a new tool for probing the circuit-level mechanisms of mental navigation and spatial imagination.

Since neuroprosthetic BMIs emerged about two decades ago (*3*), a basic paradigm has been used: Animals, which have been surgically implanted with cortical recording electrodes, engage in a sensorimotor task—for example, moving a joystick to control a cursor on a screen. Initial trials are used to train a "decoder," which maps neural activity to the task space. Once trained, the task interface is switched to be controlled by decoded ongoing neural activity rather than the animal's physical movements. This approach has provided information on the neural circuits that plan and control limb movements (*4*) as well as enabling devices that restore communication function in individuals with severe paralysis (*5*). So far, BMIs have largely targeted neurons from

This conclusion challenges the common perception that vegetation restoration often results in faunal restoration through passive re-colonization (*6*) and highlights that ecosystem disassembly and reassembly are not symmetrical. In terrestrial systems, for example, many wild animals are not likely to recolonize (relatively small) restored vegetation patches that are embedded in highly fragmented landscapes. This could be due to barriers to movement (*7*) or to local, regional, or functional extinctions (*1*, *8*).

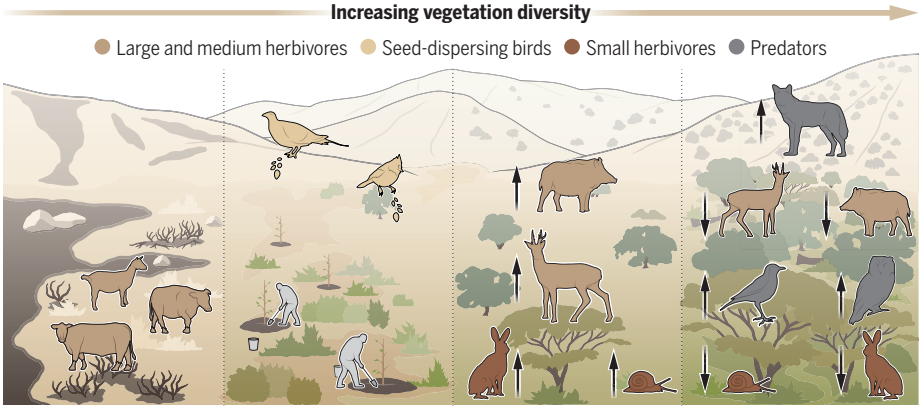
An established principle in ecology is that, through trophic and behavioral cascades, predators limit herbivore densities, which reduces consumption of vegetation and thus increases vegetation biomass (*9*). Xu *et al.* reviewed the few available studies of the impact of predator reintroductions on vegetation restoration. The results suggest that predator reintroduction might be as effective for vegetation restoration as removing herbivores (through physical exclosures, insecticides, or deterrents) or plant-centered solutions (such as decreasing plant competition or promoting facilitation), indicating that the presence of predators in an ecosystem plays a critical role in the restoration of its vegetation (see the figure). A major benefit of this nature-based solution is that it enables vegetation restoration while also restoring the trophic structure and potentially the composition and food webs of degraded ecosystems.

All predator reintroduction studies analyzed by Xu *et al.* come from temperate terrestrial ecosystems, where wildlife, in general, and predators, in particular, are slowly returning (*10*). The findings suggest that this could accelerate vegetation regeneration. In tropical and subtropical systems, predator and herbivore species whose numbers are declining (*9*) could be reintroduced for concomitant restoration of vegetation and trophic structure. Yet empirical field studies need to be conducted to validate this hypothesis.

Although the study of Xu *et al.* supports the importance of trophic cascades and the key role of trophic rewilding in ecosystem restoration, the specific restoration solutions applied should also consider local contingencies and the broader socioeconomic context. Their work also emphasizes that restoration of the diversity in Earth's ecosystems is intrinsically linked to the restoration of their biostructure. ■

Restoring food webs to regenerate vegetation

Restoration of ecosystems such as forests can be initiated by the active planting of vegetation or by natural processes such as seed dispersal by birds. However, excessive numbers of herbivores and imbalances in the composition of herbivore communities can limit regeneration through high levels of herbivory. Reintroducing predators to the ecosystem can control herbivory and facilitate restoration.



Degraded ecosystem
Vegetation is lost through fire, deforestation, and livestock overgrazing.

Restoration efforts
Planting of diverse vegetation and natural regeneration are steps toward restoration.

Excessive herbivory
Excessive numbers of herbivores consume grass and young trees.

Predation
Reintroduction of predators controls herbivory and increases trophic diversity.

GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. R. Dirzo *et al.*, *Science* **345**, 401 (2014).
2. O. Schmitz *et al.*, *Nat. Clim. Change* **13**, 324 (2023).
3. United Nations (UN), "UN Decade on Ecosystem Restoration: Action plan booklet" (UN, 2023).
4. C. Xu *et al.*, *Science* **382**, 589 (2023).
5. E. S. Bakker *et al.*, *Aquat. Bot.* **135**, 18 (2016).
6. L. Poorter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2003405118 (2021).
7. M. A. Tucker *et al.*, *Science* **359**, 466 (2018).
8. A. Benítez-López *et al.*, *Science* **356**, 180 (2017).
9. M. L. Pace, J. J. Cole, S. R. Carpenter, J. F. Kitchell, *Trends Ecol. Evol.* **14**, 483 (1999).
10. S. E. H. Ledger *et al.*, "Wildlife comeback in Europe: Opportunities and challenges for species recovery" (Rewilding Europe, 2022).

10.1126/science.adl0578

NEUROSCIENCE

The neural basis of mental navigation in rats

A brain–machine interface demonstrates volitional control of hippocampal activity

By **Michael E. Coulter**[1] *and* **Caleb Kemere**[2]

Complex behaviors such as navigation often require chains of decisions that can be mentally simulated ahead of carrying out the behavior. The activity of spatially coding neurons (place cells) in the hippocampus and associated cortical regions has been posited as the expression of a "cognitive map," which could serve as a substrate for generating these complex behaviors (*1*). On page 566 of this issue, Lai *et al.* (*2*) report that rats can navigate through a virtual reality (VR) environment using a brain–machine interface (BMI) that estimates the rat's mental location from the ongoing activity of neurons in its hippocampus. This finding is an exciting expansion of BMI implementation from sensorimotor functions to a more cognitive domain and suggests that hippocampal activity is under volitional control. In addition, the BMI approach provides a new tool for probing the circuit-level mechanisms of mental navigation and spatial imagination.

Since neuroprosthetic BMIs emerged about two decades ago (*3*), a basic paradigm has been used: Animals, which have been surgically implanted with cortical recording electrodes, engage in a sensorimotor task—for example, moving a joystick to control a cursor on a screen. Initial trials are used to train a "decoder," which maps neural activity to the task space. Once trained, the task interface is switched to be controlled by decoded ongoing neural activity rather than the animal's physical movements. This approach has provided information on the neural circuits that plan and control limb movements (*4*) as well as enabling devices that restore communication function in individuals with severe paralysis (*5*). So far, BMIs have largely targeted neurons from

regions whose activity strongly correlates with movement and so might be expected to be under natural volitional control.

There has been interest in BMIs that decode cognitive states (for example, affect or preference) for neurofeedback or closed-loop modulation (*6, 7*). Critically, unlike BMIs that use naturally volitional regions, cognitive neurofeedback requires the user to practice to achieve the desired patterns of neural activity. Although there have been some reports that hippocampal activity might precede motor actions (*8, 9*), the activity patterns of hippocampal place cells are thought to be largely generated from action by the integration of sensory information with internal feedback on limb position and movement. Whether the population code in the hippocampus could be volitionally generated naturally (without training) remained uncertain.

The BMI system developed by Lai *et al.* decodes the activity of sparsely firing hippocampal place cell ensembles. The authors used a task in which rats initially navigate through a VR environment to a visually cued reward location by running on a spherical treadmill. A decoder is trained on the ensemble neural activity recorded during this initial training period. Then, control of the visual scene is transferred from the treadmill to the signal that is decoded in real time from hippocampal activity. Unexpectedly, they found that under BMI control, the rat continues to navigate through the VR environment to reward locations. This observation provides evidence that the ensemble activity in the hippocampus can be volitionally controlled without training.

Spatial navigation can be thought of as model-based control (*10*): By using a predictive model of the world, actions are planned. Executing these actions changes the relationship of the individual to the world. This, in turn, updates sensory information, which finally updates predictions from the world model. This process is fundamentally a loop, but it critically involves motor action and sensory feedback. In addition to physically traversing the world, humans can mentally navigate—for example, choosing routes by imagining traveling through them. This imagined navigation coarsely activates the same temporal lobe brain regions as actual navigation (*11*). However, a deeper circuit-level understanding of mental navigation has been prevented by an inability to instruct a nonhuman subject to imagine a navigational path.

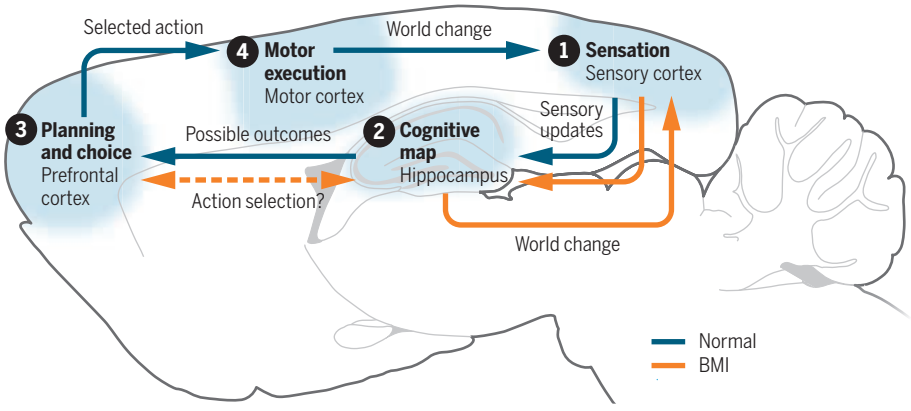
By directly tying changes in the VR environment to the output of the hippocampus, the approach taken by Lai *et al.* may enable the study of mental navigation in animal models (see the figure). Indeed, while rats primarily ran on the treadmill during BMI control, on many trials they successfully navigated the VR while remaining still. Moreover, if the decoded signal was used to move an object in a stationary three-dimensional environment, the animals were capable of directing this movement with their hippocampal activity. Unfortunately, the authors could not distinguish between the hypothesis that the animals are continuing to mentally navigate as if they were the object that was moving, as a human might look at a picture and imagine themself moving in space, or whether they were directly representing the location of a distinct object by activating recently reported hippocampal neurons associated with object locations (*12*) (which would be less akin to mental navigation).

Can animals be trained to perform imagined navigation—to volitionally activate representations of remote locations—in the real world? Previous work suggests that representations of space by the activity of hippocampal neurons are similar in VR and real environments (*13*), so it is likely that the findings of Lai *et al.* also apply outside of VR; future experiments can test this hypothesis.

The hippocampus is known to exhibit bursts of ensemble activity in sequences that seem to represent locations away from the animal's current position. These bursts take place during sharp wave ripples (SWRs), which occur during periods of rest and sleep, and theta sequences, which are synchronized by oscillations in activity present during ongoing exploration. Both types of burst sequences have been given hypothetical roles in mental navigation (*14*), but the BMI approach of Lai *et al.* presents an opportunity to test this directly. They observed periods of SWRs and theta oscillations during performance of the BMI task. Strong theta oscillations were observed during all trials, when the rat was moving and still. This latter observation is consistent with rats mentally moving through the environment. Although post hoc analyses found equivalent navigation ability when SWRs were excluded, the 1.5 to 5 s decoding windows used in the BMI were too long to assess whether theta sequences are present or how they might be used by the animal to navigate. Developing faster online decoding algorithms and/or using higher-density neural recording techniques might address this issue and enable experiments to uncover the timescale of the ensemble activity that enables mental navigation. ■

Investigating the neural basis of mental navigation

To enable spatial navigation, sensory information (**1**) is used to generate a predictive model of the world [cognitive map (**2**)], which enables actions to be planned (**3**). Execution of the planned action (**4**) changes the relationship of the individual to the world, which updates sensory information. A brain–machine interface (BMI) was trained to decode place cell activity from the hippocampus of rats navigating in virtual reality. The animals were able to control their location through the BMI, thus enabling changes to their relationship with the world in the absence of motor execution. This approach facilitates investigation of the neural circuits underlying mental navigation.



[1]University of California San Francisco, San Francisco, CA, USA. [2]Rice University, Houston, TX, USA. Email: michael.coulter@ucsf.edu; caleb.kemere@rice.edu

REFERENCES AND NOTES

1. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
2. C. Lai *et al.*, *Science* **382**, 566 (2023).
3. A. B. Schwartz, *Annu. Rev. Neurosci.* **27**, 487 (2004).
4. N. G. Hatsopoulos, J. P. Donoghue, *Annu. Rev. Neurosci.* **32**, 249 (2009).
5. F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, *Nature* **593**, 249 (2021).
6. K. A. Moxon, G. Foffani, *Neuron* **86**, 55 (2015).
7. M. M. Shanechi, *Nat. Neurosci.* **22**, 1554 (2019).
8. E. Halgren, *Hippocampus* **1**, 153 (1991).
9. J. C. Arezzo, C. E. Tenke, H. G. Vaughan Jr., *Brain Res.* **401**, 79 (1987).
10. R. J. Dolan, P. Dayan, *Neuron* **80**, 312 (2013).
11. A. J. Horner, J. A. Bisby, E. Zotow, D. Bush, N. Burgess, *Curr. Biol.* **26**, 842 (2016).
12. D. B. Omer, S. R. Maimon, L. Las, N. Ulanovsky, *Science* **359**, 218 (2018).
13. G. Chen, J. A. King, Y. Lu, F. Cacucci, N. Burgess, *eLife* **7**, e34789 (2018).
14. G. Pezzulo, C. Kemere, M. A. A. van der Meer, *Ann. N. Y. Acad. Sci.* **1396**, 144 (2017).

10.1126/science.adl0806

RETROSPECTIVE

Ferid Murad (1936–2023)

Physician-scientist who discovered NO signaling

By **Iraida Sharina** *and* **Emil Martin**

Ferid "Fred" Murad, the pharmacologist and physician who made paradigm-shifting discoveries about the molecular underpinnings of cardiovascular physiology, died on 4 September. He was 86. Murad's groundbreaking discovery of the role of nitric oxide (NO) in cell signaling revolutionized the field of gaseous second messengers and cyclic nucleotide signaling in physiology, and his work provided a scientific basis for the development of a multitude of new therapies for cardiovascular and other diseases.

Born on 14 September 1936 in Whiting, Indiana, Murad graduated in 1958 from DePauw University in Greencastle, Indiana, with BS degrees in premedical science and chemistry. That same year, he married Carol Ann Leopold, with whom he would spend the rest of his life. In 1965, he was one of the first graduates of the MD-PhD program at Case Western Reserve University. Upon completion of his residency at Massachusetts General Hospital in 1967, Murad took a job at the National Institutes of Health. In 1970, the University of Virginia recruited him to develop a new clinical pharmacology division in the Department of Medicine. In 1981, Murad became the chief of medicine at the Palo Alto Veterans Hospital in California, while serving as an acting chair of internal medicine and pharmacology at Stanford University. He later became vice president of Abbott Laboratories. After several years, feeling restricted by the industrial biotech environment, Murad returned to academia, joining the University of Texas Health Science Center in Houston. In 2010, he moved to George Washington University; then, in 2016, he came back to Stanford as an emeritus professor.

In 1975, Murad identified soluble guanylyl cyclase as a receptor and critical transducer of NO signaling in vascular smooth muscle cells. He demonstrated that NO activates guanylyl cyclase, inducing the production of another second messenger, cyclic guanosine monophosphate (cGMP), which in turn activates scores of other enzymes, further multiplying and propagating the signal. He found that nitroglycerin is converted to NO and, in turn, relaxes the smooth muscle cells that surround blood vessels. The increase in relaxation causes the vessels to widen, which facilitates blood flow. Murad found that several other commonly used vasodilators also serve as a source of NO. This discovery solved the mystery of how nitroglycerin relieves the symptoms of heart conditions, including angina and coronary artery disease. The work carried out in his laboratory established NO–cGMP signaling as a ubiquitous pathway in physiology that affects every organ system.

PHOTO: DEPAUW UNIVERSITY

The idea that NO, a free radical with a short life span, could act as a ubiquitous second messenger in a highly reactive tissue environment was originally received with skepticism by the scientific community. Murad often recalled that in the 1970s his talks at conferences were regularly scheduled as the last presentation on the final day, when most attendees had already departed, and were often met with open laughter from the audience. It is thanks to his unwavering confidence, often bordering on stubbornness, and his unstoppable devotion to his ideas that, by the end of the 1980s, the function of NO as a signaling factor in the cardiovascular system, much like known hormones or neurotransmitters, was widely accepted.

Murad's relentless scientific pursuits earned him a share of the 1998 Nobel Prize in Physiology or Medicine. His other awards included the Albert Lasker Basic Medical Research Award in 1996, the Golden Plate Award of the Amer Academy of Achievement in 1999, the Ciba Award for Hypertension Research of the American Heart Association in 1988, and the Baxter Award of the Association of American Medical Colleges in 2000. He was also an active member of various scientific academies.

We both joined Fred's laboratory in the late 1990s as postdoctoral fellows and were immediately swept into the whirlwind of his scientific ideas and boundless curiosity. He loved to puzzle his trainees with questions related to NO biology, challenging them to prove or disprove points with the available research. He collected the best minds from around the world, from countless conferences and even golf games, and added them to his big, multinational laboratory. Different cultures and diverse schools of thought intrigued him, and he fostered a supportive training environment. He loved to share his ideas with anyone willing to listen, and he took a personal interest in his students and colleagues. Fred will be remembered for his incredible contributions to medicine and science, but his ultimate scientific legacy is the hundreds of prominent researchers, honored to call him a mentor, who continue his scientific pursuits throughout the world.

Fred had an amazing, sometimes controversial, but always larger-than-life personality. He told remarkable stories, which ranged from a secret meeting with Yasser Arafat to a failed attempt to raise cows for food. To emphasize the importance of hard work and determination, he often recalled how he and Carol took extra jobs to earn the money they needed during his time in graduate school. He spoke of his independence, industrious spirit, and willingness to work with pride and humor.

Fred loved and cherished both his work and his family, and he would never agree to sacrifice one in favor of the other. When asked what he was going to do in retirement, Fred would always insist that he was planning to die at the lab, and we believed him. He felt that life, no matter how hard, is an adventure to be celebrated and enjoyed and that people are fundamentally good. His confidence and optimism were contagious. Fred was an extraordinary physician, scientist, and educator. His pioneering research continues to influence today's development of cutting-edge treatments for hypertension, heart and lung conditions, cancer, shock, and other diseases. With his death, the world lost a great, ever-inquisitive mind and a fearless advocate of the scientific approach to medicine. ■

Division of Cardiology, Department of Internal Medicine, McGovern Medical School, UTHealth Houston, Houston, TX, USA. Email: iraida.g.sharina@uth.tmc.edu

10.1126/science.adl1754

POLICY FORUM

BIODIVERSITY

New benefit-sharing principles for digital sequence information

Benefit sharing should be decoupled from access

By **Michael Halewood[1], Margo A. Bagley[2], Markus Wyss[3], Amber Hartman Scholz[4]**

Since 1992, international laws governing access and benefit sharing (ABS) have confirmed the principle that countries have sovereign rights over their nonhuman biological diversity and can regulate access to their genetic resources. These laws require the fair and equitable sharing of monetary (e.g., royalties) and nonmonetary (e.g., scientific training) benefits resulting from physical access to and utilization of those genetic resources (GRs). The Kunming-Montreal Global Biodiversity Framework, adopted by the Conference of the Parties to the Convention on Biological Diversity (CBD) at their 15th meeting (COP-15) in December 2022 (*1*), includes a decision to establish a separate, multilateral benefit-sharing mechanism for the use of "digital sequence information" (DSI), i.e., biological data associated with, or derived from, genetic resources such as nucleotide sequences and epigenetic, protein, and metabolite data. The international community has a narrow window of opportunity to develop a DSI benefit-sharing framework that is simple, harmonized, effective, and transformational.

For now, the COP-15 decision is an empty shell, with no agreement on how benefits will be shared, by whom, for what, and under what conditions. These issues will be addressed over the next few years. In parallel, other UN fora, the Food and Agriculture Organization, the World Health Organization (WHO), and the Convention on the Law of the Seas are also grappling with how to address benefit sharing from the use of DSI.

HOW ABS CURRENTLY WORKS

The concept that benefits from the utilization of GRs should be equitably shared is by now well established, and one of the primary objectives of several international agreements over the past 30 years, including the CBD (1992), its Nagoya Protocol (NP, 2010), the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty, 2001), and the Pandemic Influenza Preparedness (PIP, 2011) Framework. Notably, the CBD and NP do not require countries to make ABS laws, meaning that each country can decide whether to regulate access to its GRs or not.

The CBD and its NP embrace a "bilateral" approach to ABS, in which a user and a provider conclude an individualized agreement that must be approved by a national authority. Each country creates its own ABS system and can choose, inter alia, what type of biodiversity is regulated, what benefits must be shared, and how permits are granted. Provider countries must process each inquiry from users, often with very limited capacity and resources. This high level of complexity and granularity is challenging and time-consuming for all, leading in some cases to users (i) pivoting away from planned research (with societal losses in innovation) or (ii) identifying nonregulated (but still legal) sources of the desired GRs (with lost opportunities for benefit sharing). In the latter case, users may still have a compliance burden to prove that their access to GRs was without ABS obligations.

To address some of these challenges, the Plant Treaty developed a multilateral ABS system to virtually pool the genetic resources of 64 crops and forages used in food and agriculture, and to facilitate access for users in each other's jurisdictions. Commercial users must make royalty payments to an international benefit-sharing fund, from sales of new products derived from accessed genetic resources, but only when those products are not made available for further research and breeding. To date, few commercial users have triggered this benefit-sharing obligation and made royalty payments. Contracting parties and others have voluntarily contributed $30 million.

In response to sovereignty claims over influenza samples, the PIP Framework facilitates access to genetic resources in the WHO Global Influenza Surveillance and Response System (GISRS) by a multilateral system in which vaccine manufacturers receive access to GISRS samples and, in turn, make annual (voluntary) monetary contributions and provide low-cost seasonal influenza vaccines. As coronaviruses were not covered, the PIP Framework was not tested during the COVID-19 pandemic. The WHO Pandemic Preparedness CA+ negotiation process will attempt to broaden pathogen coverage by spring 2024.

In June 2023, Parties agreed to a new multilateral instrument, the UN High Seas Treaty, also known as the Biodiversity Beyond National Jurisdiction (BBNJ) Treaty, under which genetic resources from international waters and DSI derived therefrom will be regulated under a to-be-developed benefit-sharing mechanism. These systems reflect the conviction that, to be "fair and equitable," benefits must be shared by entities that have physically accessed, and directly used, GRs in the course of research and development (R&D). We postulate that coupling of access and benefit sharing is one of the main reasons why ABS approaches are largely ineffective, and that a new system for DSI should largely if not entirely decouple benefit sharing from access. Although the COP-15 decision is focused on DSI, and the NP will continue to apply to GRs, we believe that parties should be allowed to subject GRs to the same harmonized, multilateral conditions in the future.

NEW DESIGN PRINCIPLES

The principle of benefit sharing has helped to facilitate political compromise (*2*), raise awareness of the rights of countries and Indigenous peoples and local communities (IPLCs), and encourage public research organizations and companies to engage in responsible sourcing of genetic resources. Yet, despite decades of efforts to implement these agreements, they have not resulted in satisfactory monetary benefit sharing to support biodiversity conservation or other priorities. The CBD and NP have also not substantially increased access to GRs, if at all (*3*, *4*); the Plant Treaty (*5*) and the PIP Framework have charted some success.

A fundamental disconnect exists between the laudable policy vision of benefit sharing and its implementation. Here we summarize key problems with existing ABS systems and provide recommendations for the new mechanism for DSI.

Problem 1: High transaction costs and low levels of legal certainty

To implement the CBD or NP, many countries have embraced highly granular, individual-transaction–based approaches to ABS. This requires complex systems to, for example, evaluate access applications, track and trace uses of accessed materials in R&D and commercialization of derived products, and cre-

ate formulas for calculating payments for products derived from multiple GRs. This complexity imposes unsustainable demands and transaction costs on users, providers, and public authorities. It also frequently results in a lack of implementation, legal uncertainty, avoidance, and lost opportunities for benefit sharing.

Even under the Plant Treaty's multilateral ABS system, the monetary benefit-sharing formula requires tracking and tracing the use of accessed genetic materials from R&D to commercialization of individual products. High transaction costs, to prove or disprove when payments are due, have driven away commercial users who can (legally) avoid using the multilateral system (*6*). Under the nominally standardized PIP Framework, conclusion of a Standard Material Transfer Agreement typically involves lengthy negotiations that are incompatible with modern, fast-paced R&D. The nature of DSI makes it particularly unsuited to the kinds of controls upon which individual-transaction–based ABS regulations depend.

Recommendation 1: Radical simplification

Benefit-sharing obligations should be calculated at an aggregate level rather than triggered by access to DSI or use of DSI in product development. During negotiations leading up to COP-15, the African regional group's proposal of a 1% levy on retail sales of biodiversity-based products reflected such an aggregate approach. A similar but simpler approach would be to base payment obligations on national gross domestic product or another appropriate UN statistic (*7*). Countries that fall below an established threshold could be exempted.

In this simplified system, governments would assume an intermediary role, making initial payments to the global benefit-sharing fund. They would then collect from their domestic economy on the basis of nationally defined criteria (*8*) without regulating access to datasets, tracking use, or adjudicating on the contribution of accessed DSI on a product-by-product basis. This would allow DSI users whose data are governed by multiple UN fora to cover payment obligations all at once. We recommend that payment obligations are standardized at the international level, on the basis of commercial activity in a given country.

The design principles should be simplicity, legal certainty, predictability, and efficiency. The goals should be to increase benefit sharing and better support conservation and sustainable use of biodiversity. This will require straightforward definitions of who will be obliged to contribute to the fund. It will also solve a perpetual problem in enforcement of the NP—that authorities do not know who individual users are (or should be).

There is precedent for such an approach. The "interim" BBNJ benefit-sharing arrangement requires parties to make direct payments to an international benefit-sharing fund. Under the Plant Treaty, negotiators are considering a subscription system with payments based on sales of subscribers' entire portfolio of crops, regardless of whether they are derived from accessed materials (and possibly DSI). Although this would lower transaction costs for each subscriber, the problem of potential avoidance by entities that choose not to subscribe remains. Avoidance within parties is not possible under our proposal. However, all benefit-sharing options under consideration would suffer if major user countries do not participate, likely leading to strict national ABS legislation on DSI and decreased willingness of parties to make DSI open access.

"...DSI benefit-sharing... must be all-encompassing, unavoidable, and...simplified."

Many parties are hesitant to commit to upfront payments, or to impose generalized benefit-sharing obligations. They prefer ABS systems in which users make payments on their own, after entering into ABS agreements. But 30 years have proven that such ABS systems are fundamentally flawed. Some might object that DSI users in nonparties would not enjoy the same protection as users in contracting parties against potential ethical or legal challenges while using DSI in open access infrastructures. A remedial option would be to allow such entities to make payments based on a simple formula, thereby "purchasing" the ability to demonstrate compliance with internationally agreed benefit sharing. A similar approach could be taken for multinational companies that have branches located in both parties and nonparties. Over time, we hope the new system will have near-universal membership, reducing the number of entities subject to such measures.

As part of this approach, rules for access to DSI need to be simple. It will be important to confirm what DSI is available for use, without restrictions and with full legal certainty, by users in participating countries. Presumably, that should include DSI that is (i) currently available through open access infrastructures and (ii) generated in the future using public funds (with possible exemptions for, e.g., privacy or IPLC rights). This is already happening; most public funding agencies require publication of research data in open access repositories. The CBD DSI decision and BBNJ treaty acknowledge the importance of open access and global databases. Furthermore, contracting parties could create incentives for actors generating DSI with private funds to deposit DSI in open access databases.

Problem 2: Benefit-sharing base is too narrowly defined and avoidance too easy

Current ABS systems have left many GRs and types of use "out of scope" for historical and legal reasons, for example, GRs in most developed countries, which have generated over half of the DSI in public databases (*9*). The increasing complexity of ABS systems incentivizes many users to access GRs through "free," nonregulated sources. A DSI system that bases payments on geographical origin or a specific data type will drive users to seek other resources. Narrow scope also limits nonmonetary benefit sharing to what single users of GRs can provide in return for access. This piecemeal approach fails to provide the scale of capacity strengthening needed to conserve biodiversity and sustainably use GRs and DSI.

A related problem is that current benefit-sharing models impose payment obligations only on those engaged in a particular kind of utilization, i.e., performing commercial R&D. They miss out on benefit sharing from a broad range of downstream industries that generate revenue by using technologies developed upstream. Under the Plant Treaty, for example, seed companies are required to make payments, but not international grain merchants or food companies. In pharmaceutical development, brand-name manufacturers may need ABS permits, but generics manufacturers that copy the drug may not. Tragically, during the COVID-19 pandemic, the PIP Framework could not support equitable access to vaccines owing to narrow scope.

Recommendation 2: Broaden the base for benefit-sharing contributions

Whereas under the NP, each country is free to establish (or not) ABS obligations for access to its GRs, the DSI COP-15 decision should result in benefit-sharing obligations for all CBD parties while offering legal open access to DSI in public databases. The scope for DSI benefit-sharing obligations must be all-encompassing, unavoidable, and radically simplified. A universal scope that covers all types of biological data and all uses of—and impacts on—biological diversity in all coun-

[1]Alliance of Bioversity International and the International Center for Tropical Agriculture, Rome, Italy. [2]Emory University School of Law, Atlanta, GA, USA. [3]DSM Nutritional Products Ltd, Kaiseraugst, Switzerland. [4]Leibniz Institute DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. Email: amber.h.scholz@dsmz.de

tries is easier to understand for providers and users. A broad scope requires an aggregate approach because of the size and complexity of DSI datasets.

Some providers may hope to redress misappropriation of biodiversity by designing a DSI benefit-sharing system that will ensure direct return of benefits to their communities. However, globalization and the omics revolution render national borders increasingly porous in relation to biodiversity and DSI. With proliferating alternative sources of supply, it is becoming harder, not easier, for countries to control access to genetic material and DSI, and it would continue to be easier for users to avoid transactionally based benefit-sharing obligations by accessing DSI through unregulated sources. We thus advocate directly addressing legal avoidance by building systems that are not coupled to physical or historical access. This will ensure that benefits are shared from these globalized and historically accessed GRs, albeit in the form of payments to a global fund. DSI utilization must be defined broadly and simply to cover all DSI. Benefit-sharing rules should be harmonized in all countries participating in the multilateral mechanism, so that payment cannot be avoided.

Finally, there is no "polluter pays" principle in the CBD. Our proposal creates space to include industries and sectors whose negative externalities contribute to the erosion of biological diversity.

Problem 3: No guarantee of benefits leads to mistrust and restrictions

Dissatisfaction with monetary benefit sharing has perpetuated distrust from many countries that consider themselves to be access providers, leading some countries to further restrict access to GRs (*10*). Many countries, betting on their endemic biodiversity to bring users to the negotiating table, remain reluctant to give up on their bilateral ABS controls and embrace multilateral benefit sharing when there is no guarantee of monetary benefits being shared directly with them.

Even the Plant Treaty's multilateral ABS system is not designed to fully address this. Monies are distributed from the Treaty's international benefit-sharing fund through a system of open, competitive bids. Some countries have made repeated bids without ever being selected, leading to their reluctance to make materials available.

Recommendation 3: Guarantee benefit sharing

The new multilateral benefit-sharing mechanism for DSI needs to guarantee minimum benefits for target beneficiaries. Monies could be allocated to project-based "streams" to achieve policy priorities, for example, to support conservation in countries with high levels of biological diversity; to support IPLCs as primary custodians of biodiversity; to increase capacity in low- and middle-income countries (LMICs) to better use DSI; to recognize countries proactively giving access to GRs used to produce DSI (*11*); or to harmonize DSI approaches (across, e.g., BBNJ, WHO, Plant Treaty, CBD). There should be a guarantee that all LMICs will periodically receive support through one or more streams.

Problem 4: The growing complexity of life sciences research

The traditional narrative of ABS envisions a rare plant harnessed to develop a blockbuster drug, under a single benefit-sharing agreement between the providing country and a biotech company. Yet in the modern, integrative, systems-based process of biological research and innovation (*12*), the chances of a product being based on a single GR, let alone a single piece of DSI, are rare. Life scientists use hundreds of millions of sequence, protein, small-molecule, structural, and pathway data points to decide what hypotheses to pursue. They build consensus sequences based on thousands of DSI entries and adapt them synthetically to optimize expression or folding. Any approach that requires benefit sharing back to the providers of the relatively narrow range of genetic resources that are identified and used at the final stages of product development inequitably ignores valuable contributions of all the genetic diversity, sourced from around the world, used in the upstream stages of R&D.

Artificial intelligence (AI) is revolutionizing biology. The output of AI is based on multiple streams of biological data, none of which belongs to a single country or UN forum, making attribution of DSI-based commercial outcomes to individual countries inconceivable. AI portends a future in which GR-based DSI itself may no longer be needed for product development. Thus, benefit sharing tied to access would no longer occur to any meaningful extent (*13*). These integrative analyses and outcomes are unlikely to produce meaningful benefit sharing unless we build a radically different framework.

Recommendation 4: Embrace aggregate-level benefit sharing

The COP-15 decision recognized that for DSI a transaction-based–benefit-sharing mechanism is not realistic. A new DSI benefit-sharing mechanism should reflect the value of open access to biomolecular data from around the world. Benefit-sharing rules should be the same for all biological data and harmonized across UN fora. This will be essential to maximize user compliance, eliminate avoidance, increase the sharing of benefits, and minimize transaction costs. Scientific transparency and integrity should continue to improve and make provenance information and data contributions clear. Data labels for IPLCs and metrics that show providers where data derived from their GRs has landed in public datasets will build trust and improve metadata.

SEIZE THE MOMENT

The recommendations above would represent major changes in established negotiating positions. They require relaxation of control over GRs and DSI on one hand, and a major expansion of scope and financial commitment on the other. Hard choices need to be made between simplicity and complexity, and national governments being passive, or proactively fulfilling their required role to make a global benefit-sharing system truly effective.

Countries should also be given the opportunity to voluntarily include GRs in the scope of the new multilateral mechanism. This would contribute to the harmonization and simplification of ABS that are crucial to making it truly effective and impactful. In effect, a policy experiment will be conducted in which the new DSI benefit-sharing mechanism runs in parallel with the bilateral approach for GRs under the CBD and NP. Countries can observe, measure, and determine which system generates more benefits and conserves more biodiversity and, over time, set new priorities. Through these large-scale compromises, benefit sharing can make a meaningful contribution to the conservation and sustainable use of biodiversity. ■

REFERENCES AND NOTES

1. CBD, COP-15 Decision 15/9, Digital sequence information on genetic resources (2022).
2. K. Ten Kate, S. A. Laird, *Int. Aff.* **76**, 241 (2000).
3. S. Laird *et al.*, *Science* **367**, 1200 (2020).
4. F. Humphries, M. Rourke, T. Berry, E. Englezos, C. Lawson, *J. Law Med.* **28**, 684 (2021).
5. M. Halewood *et al.*, *Plants* **9**, 1296 (2020).
6. International Seed Federation (2015); https://www.fao.org/3/ca4962en/ca4962en.pdf.
7. P. Oldham, C. Chiarolla, S. Thambisetty, LSE Law—Policy Briefing Paper no. 53 (2023); https://doi.org/10.2139/ssrn.4343130.
8. P. Oldham, J. Kindness, Sharing digital sequence information, Zenodo (2022); https://doi.org/10.5281/zenodo.6557191.
9. F. Rohden *et al.*, CBD 2020; https://www.cbd.int/doc/c/1f8f/d793/57cb114ca40cb6468f479584/dsi-ahteg-2020-01-04-en.pdf.
10. A. Sirakaya, *Front. Plant Sci.* **10**, 1175 (2019).
11. A. H. Scholz *et al.*, *Nat. Commun.* **13**, 1086 (2022).
12. F. Michiels *et al.*, *Sustainability* **14**, 277 (2022).
13. S. L. Lisanza *et al.*, *bioRxiv* 2023.05.08.539766 (2023).

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) WiLDSI VorweRts 16LW0062K (A.H.S.), Horizon Europe EVA-GLOBAL 871029 (A.H.S.), and the CGIAR Genebank Initiative (M.H.).

10.1126/science.adj1331

BOOKS *et al.*

Autonomy, dignity, equality, and community should be prioritized in technology design.

SCIENCE AND SOCIETY

Human values in a digital age

Big Tech must center human rights in data decisions, argues a political scientist

By **Nita Farahany**

Wendy Wong's *We, the Data: Human Rights in the Digital Age* is a timely and engaging exploration of the human rights implications of "datafication"—the recording, analyzing, and indefinite storing of data about individuals. Wong aligns with Shoshana Zuboff's stance in *The Age of Surveillance Capitalism*, addressing the consequences of exploiting human experience as data used to predict and modify human behavior for profit. Like Zuboff, Wong understands that this process deeply affects human experience and touches on vital human rights, including autonomy, dignity, and privacy. But *We, the Data* presents a refreshing and nuanced view of data and a recentering on the basis and motivation for human rights—autonomy, dignity, equality, and community—as the framework for how to apply existing human rights to the information age.

Wong challenges the perception of data as mere raw material, suggesting that data are more multifaceted and collective than Zuboff proposed. Instead of agreeing that individuals are relegated to being data subjects in this process, she emphasizes their role as data cocreators. This perspective shift forces us to reconsider our approach to the ubiquity of data.

Each of the book's eight chapters has a dual aim: to illustrate the transformation of the human experience in the era of pervasive data and to emphasize where autonomy, dignity, equality, and community can be influential in shaping discourse and policies on data governance. Tackling topics such as the right to be forgotten, facial recognition technology, digital immortality, and data literacy, Wong maintains a narrative thread using hypothetical scenarios, real-world examples, and personal anecdotes. Her insights are bolstered by richly detailed references to key arguments and thinkers from disciplines from diverse fields, including law, sociology, philosophy, and computer science.

A pivotal chapter, "Big Tech and Us," dissects the responsibilities of Big Tech conglomerates such as Meta (formerly Facebook), Google, and Amazon. Here, Wong rebuffs the conventional view that these companies are merely private actors that provide content and access to the internet, arguing instead that "Big Tech governs" and exercises authority and influence over social and political life through its platforms and policies.

Drawing from political science orthodoxy, deliberative democracy theory, and the concept of infrastructure, Wong shows how Big Tech platforms affect our rights, interests, and values in multiple and often hidden ways. With this framing, she convincingly argues that we should hold these corporations accountable to the ethos and full spectrum of human rights, not just to singular issues such as freedom of expression or privacy, and advocates for more transparency, participation, and oversight in their decision-making processes. She also offers practical suggestions and examples of how we can and should challenge the status quo, contrasting reactive measures like the Ranking Digital Rights project against proactive civil society measures that could enable us to reframe conversations about technology from a narrow focus on privacy and surveillance to the broader set of human rights values. As Wong puts it, "Our very humanity is at stake here."

The book culminates in a call to action focused on recognizing a human right to data literacy, which Wong defines as the ability to understand, interpret, and use data in a datafied world. Data literacy, she claims, is imperative to flourishing in the digital era and will empower us to demand more and better from companies and governments that make, disseminate, and monetize data about people. It will enable us to cocreate data and shape the data process in ways that reflect and protect our human rights values.

For readers who were expecting more-radical or comprehensive solutions, this may seem a surprising and perhaps disappointing note to end on. Wong herself acknowledges that data literacy is not a panacea and that it faces many obstacles and limitations, such as unequal access, biased algorithms, and entrenched interests. Economic, cultural, and social rights also suffer from enforceability challenges and lack the same international support as the political and civil rights in the United Nations' Universal Declaration of Human Rights.

We, the Data: Human Rights in the Digital Age
Wendy H. Wong
MIT Press, 2023. 280 pp.

But perhaps Wong's restraint is her wisdom. In the end, she has proposed a workable mechanism by which we can become informed enough to reset the terms by which data will animate our lives. Human rights are not static or fixed, but rather dynamic and evolving, and they depend on our collective action and imagination to make them meaningful and effective. Her proposal, then, is a welcome invitation for each of us to reclaim our agency and become full participants in conversations and decisions about how to apply human rights and human values to our data reality. ■

The reviewer is the Robinson O. Everett Distinguished Professor of Law and Philosophy at Duke University, Durham, NC, USA, and the author of *The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology* (St. Martin's Press, 2023).
Email: farahany@duke.edu

10.1126/science.adk4912

PHOTO: GREMLIN/ISTOCKPHOTO

HUMAN HEALTH

Rethinking reproductive harm

Fears about threats to fetal health have long fallen disproportionately on expectant mothers

By **Tara Mulder**

Policing Pregnant Bodies: From Ancient Greece to Post-Roe America
Kathleen M. Crowther
Johns Hopkins University Press, 2023. 288 pp.

A plastic fetus floats in a water-filled jar. A doll head sporting a lit cigarette between puckered lips sits on top. As the doll "smokes," tar and nicotine tinge the water dingy brown. Throughout the 20th century, this educational doll dubbed "Smokey Sue Smokes for Two" was used to demonstrate the dangers of smoking while pregnant. A millennium earlier, a Persian physician painted a similarly chilling picture of the dangers of toxic menstrual blood. Infants who retained this poisonous substance after birth would develop smallpox, he warned. Today, scientific studies have shown that cigarette smoke harms fetuses and disproven any connection between menstrual blood and smallpox, but the same unspoken assumption is nonetheless present in both cases: The womb is a dangerous place for the developing fetus. The premise of each of these warnings, argues Kathleen Crowther, is that "the major threats to the health and well-being of fetuses come from their own mothers."

In her new book, *Policing Pregnant Bodies: From Ancient Greece to Post-Roe America*, Crowther explains that "although our current biomedical understanding of pregnancy and fetal development is relatively new…many of our most cherished ideas about reproduction trace their origins back thousands of years." Crowther believes that we must examine the long history of ideas about pregnant people, fetuses, pregnancy, and abortion to understand where we are now. She does not trace a direct line of influence from the ancient world to today—an impossible task—but she effectively shows how the same misogynistic ideas crop up repeatedly throughout history, pitting pregnant people against fetuses in a dangerous zero-sum game.

Our preoccupation with fetal heartbeat detection has roots in ancient ideas that centered the heart as a key feature of humanity, maintains a historian.

The way that we have studied fetal development has long had the effect of separating fetuses from the people within whom they grow. Crowther introduces readers to ancient Greek doctors examining miscarried embryos and trying to pin down when we become human; to the macabre collection of 17th- and 18th-century anatomist Frederik Ruysch, who kept hundreds of preserved fetuses in glass jars; and to Lennart Nilsson, who photographed detached, dead fetuses for his 1965 book, *A Child is Born*. Focusing on fetal anatomy and development, she argues, obscures the fetus's "connection to and impact on the maternal body" with "disastrous implications for the health of both mothers and babies."

Since the overturning of *Roe* v. *Wade*, a number of US states have passed heartbeat bills, outlawing abortion from the first detection of cardiac activity. Pro-life activists present the fetal "heartbeat"—really an electrical artifact translated into sound by an ultrasound machine—as the "voice" of the fetus. But why this focus on the heartbeat at the expense of other markers of life, such as brain activity or breath? That is a question, Crowther argues, that must be answered by history, not science.

For thousands of years, religious, philosophical, and medical doctrine promoted the idea that the heart is what makes us human. Ancient Greek physicians and philosophers, for example, connected the heart with the soul. Starting in the 20th century, Crowther writes, obstetricians strapped electronic fetal monitors to laboring women to "communicate with the fetus in a way that the mother could not." And in 2011, a pro-life activist group arranged for two fetuses to "testify" before Ohio lawmakers: Pregnant women were placed out of view, while live ultrasound images of their 9- and 15-week-old fetuses were projected onto a screen, and the fetal "heartbeats" were amplified over the chamber speakers. (Ironically, signals from the 9-week-old fetus could not be heard over sounds coming from its mother's body.)

In *Policing Pregnant Bodies*, Crowther combines three decades of experience as a medical historian with a rare ability to communicate clearly and engagingly with a general audience. She also brings her personal experience of living in Oklahoma—"one of the reddest states in the country"—to bear. In 2021, Oklahoma passed House Bill 2441, banning abortion after 6 weeks of pregnancy. When women in the state miscarry and seek medical attention, they can be arrested for manslaughter. But Oklahoma also has some of the worst infant and maternal mortality rates in the country. Fifty-four of its 77 counties contain food deserts. Active fracking sites in the state pollute the water supply, substantially increasing the risk of miscarriage. Policies such as House Bill 2441, Crowther writes, "treat pregnant people themselves as the greatest threat to fetal life," while doing nothing to address factors that play a far greater role in infant and maternal mortality, including lack of access to healthy food, clean water, safe housing, and medical care. ■

The reviewer is at the Department of Ancient Mediterranean and Near Eastern Studies, University of British Columbia, Vancouver, BC, Canada. Email: tara.mulder@ubc.ca

10.1126/science.adk6908

PHOTO: JEFF PACHOUD/AFP VIA GETTY IMAGES

Edited by **Jennifer Sills**

Chemical simplification and tracking in plastics

The zero draft of the Global Plastics Treaty (*1*) affirms that the presence of hazardous chemicals poses a substantial barrier to the safe and environmentally sound management of plastics. Additives, nonintentionally added substances (NIASes), and contaminants in plastics complicate their reuse and disposal (*2*) and hinder plastic recycling (*3*), which has been touted as a key solution to the plastics pollution crisis (*4*). Studies show evidence of accumulation of hazardous substances even in relatively close-loop plastic recycling systems, such as those for food-grade polyethylene terephthalate (*5*), but recyclers lack the tools and information needed to identify these chemicals in waste streams, or to isolate and remove them from plastic products.

Hazardous chemicals present risks to recycling workers and consumers, as well as to the wider society and environment (*6*). However, current regulations do not require plastics producers to track or make available information on the levels of harmful chemicals (*2*). Because upstream producers lack the incentives to disclose this information, recyclers are unable to control hazardous substances.

Before recycling can contribute to tackling the plastics pollution crisis, the plastics industry must limit hazardous chemicals. International regulations and innovations can drive "chemical simplification"—i.e., the reduction of the extensive number of chemicals used in plastics production (*7*). Chemical additives known to cause harm to human health and the environment must be rapidly phased out, and NIASes must be identified and limited. To ensure these changes, plastics producers should transparently report chemicals production and use, which should be tracked and monitored throughout the plastics life cycle. Finally, strict regulations specifying where recycled plastics can be used should be implemented to prevent harmful chemicals in, for example, toys and food packaging. Including obligations to increase the safety, transparency, and traceability of the components in plastics in the plastics treaty represents one of many necessary steps toward safer recycling.

SUBMIT NOW

NEXTGEN VOICES: CHANGING EXPECTATIONS

Add your voice to *Science*! Our new NextGen Voices survey is now open: In this week's Working Life, "Putting down roots" (https://scim.ag/WL_locationmatters), S. Hejmadi questions the expectation that scientists should switch institutions between a PhD and postdoc. What traditional career expectation in your field seems outdated or exclusionary? How should that expectation be adjusted to make your field more nimble or welcoming?

To submit, go to www.science.org/nextgen-voices

Deadline for submissions is 17 November. A selection of the best responses will be published in an upcoming issue of *Science*. Submissions should be no more than 150 words. Anonymous submissions will not be considered.

Hazardous chemicals in plastics pose challenges to ensuring that recycled materials are safe.

Bethanie Carney Almroth[1]*, Tridibesh Dey[2], Therese Karlsson[3], Mengjiao Wang[4]
[1]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden. [2]Department of Global Studies, School of Culture and Society, Aarhus University, Aarhus, Denmark. [3]International Pollutants Elimination Network, Gothenburg, Sweden. [4]Greenpeace Research Laboratories, Department of Bioscience, University of Exeter, Exeter, UK.
*Corresponding authors.
Email: bethanie.carney@bioenv.gu.se

REFERENCES AND NOTES

1. United Nations Environment Programme, "Zero draft text of the international legally binding instrument on plastic pollution, including in the marine environment" (2023); https://wedocs.unep.org/handle/20.500.11822/43239.
2. United Nations Environment Programme, "Chemicals in plastics: A technical report" (2023).
3. B. D. Vogt *et al.*, *ACS Appl. Polym. Mater.* **3**, 4325 (2021).
4. A. Mah, *Plastic Unlimited: How Corporations Are Fuelling the Ecological Crisis and What We Can Do About It* (John Wiley & Sons, 2022).
5. S. Gerassimidou *et al.*, *J. Hazard. Mater.* **430**, 128410 (2022).
6. H. Yang, M. Ma, J. R. Thompson, R. J. Flower, *J. Epidemiol. Commun. Health* **72**, 237 (2018).
7. K. Fenner, M. Scheringer, *Environ. Sci. Technol.* **55**, 14470 (2023).

COMPETING INTERESTS

B.C.A. is an unpaid steering committee member of the Scientists' Coalition for an Effective Plastics Treaty. M.W. is an unpaid member of the Hazardous Substances Advisory Committee.

10.1126/science.adk9846

Plastics treaty text must center ecosystems

Plastics pollution is now ubiquitous (*1*, *2*) and affects the environment at all stages of the plastics life cycle. Plastics manufacturing involves thousands of chemicals that can pollute ecosystems. Once in the environment, plastic litter can entangle and choke wildlife, and plastic objects can act as a reservoir for invasive species and concentrate other pollutants (*1*, *3*, *4*). Plastics can then break down into potentially toxic micro- and nanoplastics. Negotiations for the plastics treaty (the legally binding instrument on plastic pollution, including in the marine environment) (*5*, *6*) must ensure that its text centers the effects of plastics on ecosystems, as services that ecosystems provide are essential for biodiversity and human health and well-being.

Plastic does not occur in isolation. Terms in the treaty's current draft such as "hotspot," "accumulation zone," "cleanup," and "sectors" focus on the distribution and concentration of pollutants rather than the natural systems these pollutants affect. Ignoring impacted ecosystems implies that

mitigating plastics pollution can take place without ecological restoration.

The current terms should be replaced by language that centers ecosystems, such as "habitats polluted by plastic products" or "ecosystems affected by plastic-associated pollution." This phrasing makes the treaty's key goals clear and tangible: Ecosystems and biodiversity must be conserved and protected from pollutants at all stages of the plastics life cycle. This language also acknowledges that each affected ecosystems will require an individualized approach to restoration, depending on where the pollution is generated, the drivers of the pollution, and the impacts to local habitats and biodiversity. Prioritizing increased ecosystem stability, protection, and restoration will be more effective long term than limited, and possibly dangerous (*7*), stopgaps.

To reduce plastics' impact on the environment, the underlying drivers of ecosystem degradation through the plastics life cycle must be addressed (*8*). Such drivers include large-scale production and consumption patterns, facilitated by fossil fuel subsidies (*9*), a lack of reuse and safe recycling options, and destructive disposal practices. Through an effective, legally binding agreement, the plastics treaty can help to implement the commitments of the Kunming-Montreal Global Biodiversity Framework (particularly targets 6, 7, 14, 16) and the Sustainable Development Goals (particularly goals 12, 14, 15). The plastics treaty should include obligations to decrease global plastics production, eliminate hazardous substances, and develop safe and sustainable plastic consumption (*10*). It should also include a timeline and roadmap for making the required changes. Resources should be allocated to each aspect of this transformation and adhere to the principle of zero waste hierarchy (*11*), which prioritizes pollution prevention over waste management measures.

Ina Tessnow-von Wysocki[1], Mengjiao Wang[2], Carmen Morales-Caselles[3], Lucy C. Woodall[4,5], Kristian Syberg[6], Bethanie Carney Almroth[7], Marina Fernandez[8], Laura Monclús[9], Scott P. Wilson[10], Matt Warren[11], Doris Knoblauch[12], Rebecca R. Helm[13]*

[1]University of Vienna, Vienna, Austria. [2]Greenpeace Research Laboratories, Department of Bioscience, University of Exeter, Exeter, UK. [3]Department of Biology, University of Cadiz, Instituto Universitario de Investigación Marina, Puerto Real, Cádiz, Spain. [4]Centre for Ecology and Conservation, University of Exeter, Penryn, UK. [5]Deep Ocean Stewardship Initiative, University of Southampton, Southampton, UK. [6]Department of Science and Environment, Roskilde University, Roskilde, Denmark. [7]Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden. [8]Instituto de Biología y Medicina Experimental, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina. [9]Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway. [10]Earthwatch Institute Australia, Carlton, VIC, Australia. [11]Plastic Free Future, Pacifica, CA, USA. [12]Ecologic Institute, Berlin. [13]Georgetown University Earth Commons Institute, Washington, DC, USA.
*Corresponding author. Email: rrhelm@gmail.com

REFERENCES AND NOTES

1. A. R. Aves *et al.*, *Cryosphere* **16**, 2127 (2022).
2. M. L. Taylor *et al.*, *Sci. Rep.* **6**, 33997 (2016).
3. S. Rist, N. B. Hartmann, in *Freshwater Microplastics: Emerging Environmental Contaminants?*, M. Wagner, S. Lambert, Eds. (Springer, Cham, 2018), pp. 25–49.
4. C. M. Rochman *et al.*, *Sci. Rep.* **5**, 14340 (2015).
5. United Nations Environment Programme, "UNEA Resolution 5/14 entitled 'End plastic pollution: Towards an international legally binding instrument'" (2022).
6. United Nations Environment Programme, "Zero draft text of the international legally binding instrument on plastic pollution, including in the marine environment" (2023); https://wedocs.unep.org/handle/20.500.11822/43239.
7. M. Spencer *et al.*, *PeerJ* **11**, e15021 (2023).
8. IPBES, "Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES)," E. S. Brondizio, J. Settele, S. Díaz, H. T. Ngo, Eds. (IPBES Secretariat, Bonn, Germany, 2019).
9. J. P. Tilsted *et al.*, *One Earth* **6**, 607 (2023).
10. N. Simon *et al.*, *Science* **373**, 47 (2021).
11. Zero Waste International Alliance, "Zero waste hierarchy of highest and best use 8.0" (2023); https://zwia.org/zwh/.

COMPETING INTERESTS

M.Wang is an unpaid member of the Hazardous Substances Advisory Committee. B.C.A., M.F., and R.R.H. are unpaid steering committee members of the Scientists' Coalition for an Effective Plastics Treaty. M.F. is an unpaid member of the Endocrine Society's delegation to the Intergovernmental Negotiating Committee on Plastic Pollution. S.W. is affiliated with the Australian Microplastic Assessment Project. R.R.H. is a paid scientific consultant for the nonprofit OceanCare.

10.1126/science.adl3202

Finance plastics reuse, redesign, and reduction

Negotiations on a new international agreement to end plastic pollution are in progress, with the next committee meeting expected to focus on the list of control measures proposed in the initial "zero" draft (*1*). The document rightly urges nations to channel financial resources toward environmentally sustainable projects. However, this treaty draft disproportionately emphasizes waste management investment and neglects opportunities for reduction, redesign, and reuse.

Supporting improved recycling is a positive but insufficient step to address plastic pollution, given that only about 10% of plastics are recycled (*2*). The majority of plastic waste is deposited in landfills (50%) or incinerated (20%), with the remainder disposed of in uncontrolled ways or leaking to the environment (*2*). Beyond recycling, essential strategies include reducing aggregate plastics production and consumption and transitioning to reuse systems (*3*). Achieving these goals will require tackling the fossil fuel–entwined political economy of plastics, which currently incentivizes investments that secure the future growth of plastic production and consumption (*4*).

Only 4% of investment capital for circularity is directed to reuse solutions, whereas recovery and recycling receive 88%, with major industry players investing heavily, especially in chemical recycling (*5*). This imbalance threatens to escalate the problem of growing volume of plastic waste (*6*). For a just and effective transition, the treaty's provisions must prioritize financial investment in other interventions. The treaty should require polymer producers to address plastic pollution by paying a substantial fee pegged to the quantity of primary plastics produced. This mechanism would disincentivize growth in fossil-based production. The treaty should also define criteria for strong and independent Extended Producer Responsibility (EPR) schemes, establishing requirements for the design, manufacturing, and management through end of life for plastic products (*7*). Governments should then transparently implement the obligations, which would incentivize private entities to invest in upstream and midstream solutions for sustainable product design, management, and reuse. As negotiations progress, the parties involved should foster a shared understanding of the imperatives to finance systems and services in line with zero draft hierarchy, including those that reduce consumption and production, encourage product redesign, and support reuse solutions. These priorities will guide the development of an effective plastics treaty.

Mengjiao Wang[1], Fredric Bauer[2]*, Kristian Syberg[3], Tom Gammage[4]

[1]Greenpeace Research Laboratories, Department of Bioscience, University of Exeter, Exeter, UK. [2]Department of Technology and Society, Lund University, Lund, Sweden. [3]Department of Science and Environment, Roskilde University, Roskilde, Denmark. [4]Center for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, Australia.
*Corresponding author.
Email: fredric.bauer@miljo.lth.se

REFERENCES AND NOTES

1. United Nations Environment Programme, "Zero draft text of the international legally binding instrument on plastic pollution, including in the marine environment" (2023); https://wedocs.unep.org/handle/20.500.11822/43239.
2. Organisation for Economic Cooperation and Development, "Global plastics outlook: Economic drivers, environmental impacts and policy options" (2022).
3. W. W. Y. Lau *et al.*, *Science* **369**, 1455 (2020).
4. J. P. Tilsted, F. Bauer, C. Deere Birkbeck, J. Skovgaard, J. Rootzén, *One Earth* **6**, 607 (2023).
5. A. Mah, *Glob. Environ. Polit.* **21**, 121 (2021).
6. K. Syberg, *Nature* **611**, S6 (2022).
7. K. Tumu *et al.*, *J. Environ. Manage.* **348**, 119242 (2023).

COMPETING INTERESTS

M.W. is an unpaid member of the Hazardous Substances Advisory Committee.

10.1126/science.adl4491

PRIZE ESSAY

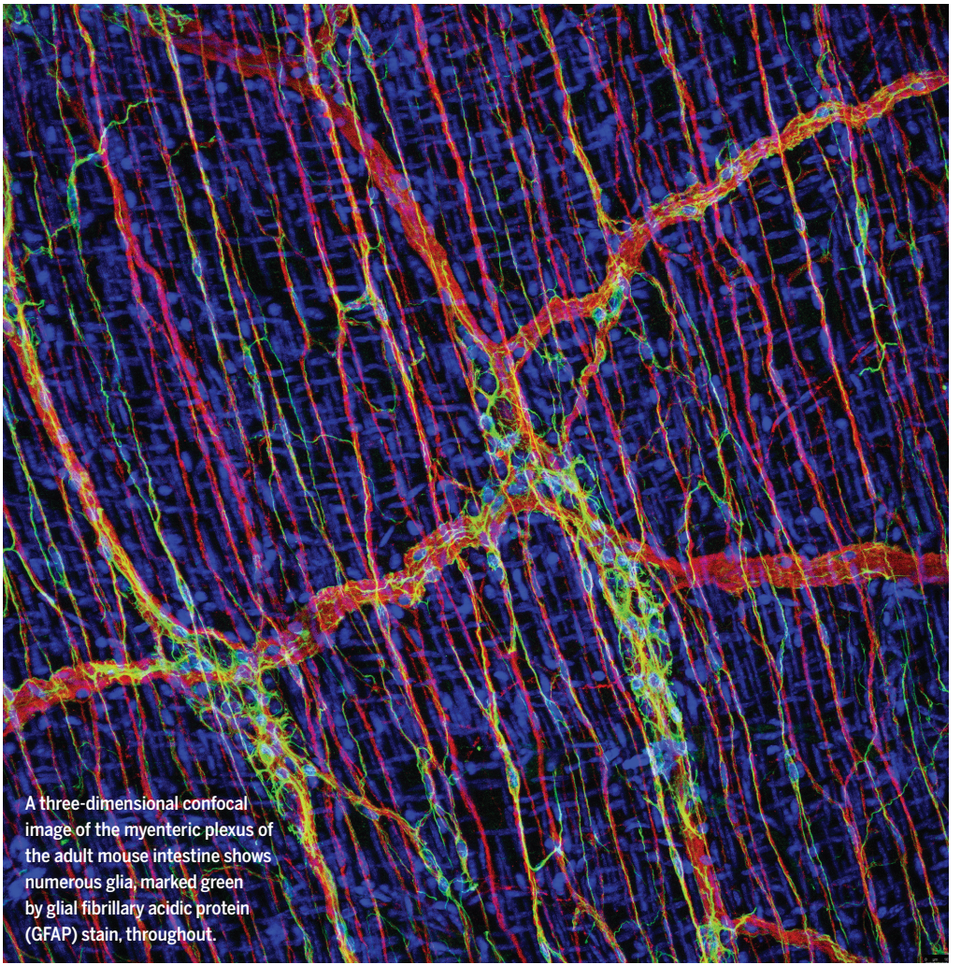
GRAND PRIZE WINNER

Marissa Scavuzzo

Marissa Scavuzzo received undergraduate degrees from Baldwin Wallace University and a PhD from Baylor College of Medicine. After completing her doctorate, she started her postdoctoral fellowship at Case Western Reserve University School of Medicine in the lab of Paul Tesar. Her research focuses on gaining a mechanistic understanding into the functional states of enteric glia in health and disease.

www.science.org/doi/10.1126/science.adk3883

eppendorf & Science PRIZE FOR NEURO BIOLOGY



A three-dimensional confocal image of the myenteric plexus of the adult mouse intestine shows numerous glia, marked green by glial fibrillary acidic protein (GFAP) stain, throughout.

PHOTOS: (LEFT TO RIGHT) JESSE ZHAN; MARISSA A. SCAVUZZO

NEUROSCIENCE

The way you move

Glial cells in the gut are specialized to fine-tune intestinal function

By **Marissa Scavuzzo**

After reading this paragraph, close your eyes and take a large drink of water. Pay attention to everything you feel as the water moves into your esophagus and down your throat. It takes conscious control and effort to move the water as you swallow. Then you can feel—but no longer control—the peristaltic wave in your esophagus as this organ takes over the action you consciously initiated. But soon after, everything you feel fades. It is not until fecal matter forms in your distal colon that you can again sense what is present and regain control of how it is moved.

Extrinsic innervation of the esophagus, stomach, and distal colon by peripheral neurons provides communication from the gut to the brain to allow you to feel and regulate these processes (*1, 2*). Everything in between is a different story. Believe it or not, your small intestine has a brain of its own that takes over for you. The "brain inside your gut," or the enteric nervous system (ENS), is a complex network of neurons and glia often referred to as the "second brain" owing to its cellular complexity and independence from central control. This second

Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. Email: marissa.scavuzzo@case.edu

1103PrizeEssay_18095533.indd 527 10/27/23 4:53 PM

brain senses the movement of contents and responds by initiating contractions to move luminal contents in a unidirectional, aboral motion independent from your brain.

In a demonstration of the gut's autonomy, when the outermost layer of the mouse small intestine is dissected and placed in culture for several days, the tissue continues to undergo peristaltic contractions without epithelial cells and any input from the central or peripheral nervous systems. This outermost layer of tissue contains smooth muscle cells, the pacemaker interstitial cells of Cajal (ICC), and the neurons and glia of the ENS. It is my goal to map the functional diversity of enteric glia in health and in a range of neurological and gastrointestinal diseases.

In 1856, Rudolf Virchow named a new cell type in the central nervous system glia, the Greek word for glue, as these cells were thought to simply hold neurons together. Glial biology is now a burgeoning field of its own, with a multitude of glial cell types identified to fulfill some of the most fundamental roles in the maintenance of brain function and integrity. Despite being a major constituent of the digestive system, the field of enteric glial biology is still in its infancy. The work that has been done in the field of enteric glial biology shows that these cells are diverse (*3*–*5*), yet their identities have not been characterized. Ablation of enteric glia leads to defects in gut motility and peristalsis (*6*–*8*), but if we only study these cells en masse, unaware of their individual roles, how will we ever reveal how this is happening?

Our laboratory first developed new cellular and molecular technologies that allowed us to define enteric glial diversity at the molecular, spatial, morphological, and tissue level (*9*). In my previous work, I found that endocrine progenitors in the pancreas are heterogeneous (*10*); in fact, there is a subpopulation of β cells called pancreatic hub cells that are less differentiated than other β cells and act as pacemakers to synchronize insulin oscillations (*11*). We noticed a specific subpopulation of enteric glia that exclusively lived in the muscle layer with distinct molecular and morphological properties (see the figure). Interestingly, these cells exhibited molecular features of less-differentiated glial cells. We named this new cell type enteric glial hub cells.

We mapped the connectome of enteric glial hub cells and found evidence that these cells interact with ICCs and enteric neurons. ICCs generate slow waves through the cyclic opening of ion channels and fluctuations in membrane potential that on their own do not cause smooth muscle contraction in the small intestine (*12, 13*). These rhythmic slow waves poise tissue for action and provide a framework for additional electrical activity to generate contractions. Thus, the gut requires external cues to reach threshold and an interplay of cell types to keep contents moving.

We found that enteric glial hub cells express Piezo-type mechanosensitive ion channel component 2 (PIEZO2), an ion channel that opens upon application of force to a cellular membrane and was a focus of the 2021 Nobel Prize in physiology or medicine (*14, 15*). We hypothesized that enteric glial hub cells respond to force to orchestrate and fine-tune the function of pacemaker ICCs and enteric neurons to coordinate muscle contractions.

When we saw that the outside layer of intestinal tissue underwent peristalsis after several days in culture, we noticed that contractions increased in response to the movement of the cell culture plate from the incubator to the microscope. Specifically, the movement and flow of media on top of the tissue seemed to be the stimuli that initiated peristalsis. Returning to the ex vivo tissue, we developed an organ-on-a-chip platform to directly apply force to the intestine while concurrently doing high-resolution calcium imaging (*9*). In this model, we found that hub cells directly sense force through PIEZO2 and observed that selective deletion of PIEZO2 from adult enteric glia in vivo led to slowing of gut motility and a buildup of contents in the stomach. When enteric glial hub cells lost the ability to sense force, the phenotype resembled gastroparesis, a poorly characterized condition in which patients are unable to empty gastric contents in a regulated manner. We also documented arrhythmic oscillations in muscle and disorganized neurotransmission in the intestinal muscle after loss of PIEZO2 from enteric glia.

These findings put enteric glial hub cells first in line to receive and sense gut contents as they regulate the level of electrical activity, which causes ICCs and enteric neurons to fire and muscle to contract. Our findings revealed that the external cue is force itself, and that these specialized enteric glia spatially position themselves by intertwining through muscle and neurons to fine-tune rhythmic contractions without any need for the brain itself. After swallowing, enteric glial hub cells in the "brain inside your gut" take over for you to keep things moving forward without it ever having to cross your mind. ■

FINALIST

Mattia Aime

Mattia Aime received his undergraduate degree in Neurobiology from the University of Pavia and a PhD in Neuroscience at the University of Bordeaux. He is currently completing his postdoctoral fellowship at the University of Bern, in the group headed by Prof. Adamantidis. His ongoing research is centered on exploring the mechanisms behind emotional memory consolidation during sleep, with the goal of identifying potential new therapeutic targets for treating maladaptive processing of emotional memories and their early sleep-dependent consolidation.

www.science.org/doi/10.1126/science.adk3894

FINALIST

Michael A. Skinnider

Michael Skinnider received his undergraduate degree from McMaster University and an MD/PhD from the University of British Columbia, during which time he was also a visiting PhD student at the École Polytechnique Fédérale de Lausanne. He then started his laboratory in the Lewis-Sigler Institute for Integrative Genomics and the Ludwig Institute for Cancer Research at Princeton University. His research focuses on the application of machine learning to problems in biology, chemistry, and medicine.

www.science.org/doi/10.1126/science.adk3912

REFERENCES AND NOTES

1. M. R. Servin-Vences *et al.*, *Cell* **186**, 3386 (2023).
2. J. Kitagawa, T. Shingai, Y. Takahashi, Y. Yamada, *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **282**, R1342 (2002).
3. L. Seguella, J. L. McClain, G. Esposito, B. D. Gulbransen, *J. Neurosci.* **42**, 8694 (2022).
4. W. Boesmans, R. Lasrado, P. Vanden Berghe, V. Pachnis, *Glia* **63**, 229 (2015).
5. M. B. Baghdadi *et al.*, *Cell Stem Cell* **29**, 86 (2022).
6. J. McClain *et al.*, *Gastroenterology* **146**, 497 (2014).
7. V. Grubišić, B. D. Gulbransen, *J. Physiol.* **595**, 3409 (2017).
8. M. Rao *et al.*, *Gastroenterology* **153**, 1068 (2017).
9. M. A. Scavuzzo *et al.*, bioRxiv 2023.2006.2007.544052 (2023).
10. M. A. Scavuzzo *et al.*, *Nat. Commun.* **9**, 3356 (2018).
11. N. R. Johnston *et al.*, *Cell Metab.* **24**, 389 (2016).
12. S. R. Cajal, *Maloine (Paris)* **2**, 891 (1911).
13. N. E. Diamant, A. Bortoff, *Am. J. Physiol.* **216**, 301 (1969).
14. B. Coste *et al.*, *Science* **330**, 55 (2010).
15. D. M. Bautista *et al.*, *Nature* **448**, 204 (2007).

ACKNOWLEDGMENTS

Past and present support for M.S. has been provided by a Howard Hughes Medical Institute Hanna H. Gray Fellowship, a New York Stem Cell Foundation Druckenmiller Fellowship, and a Simons Foundation Autism Research Initiative Pilot Award.

10.1126/science.adk3883

PHOTOS: (TOP TO BOTTOM) COURTESY OF MATTIA AIME; SAMEER KHAN

1103PrizeEssay_18095533.indd 528 10/27/23 4:53 PM

PRIZE ESSAY

FINALIST
Mattia Aime

Mattia Aime received his undergraduate degree in Neurobiology from the University of Pavia and a PhD in Neuroscience at the University of Bordeaux. He is currently completing his postdoctoral fellowship at the University of Bern, in the group headed by A. Adamantidis. His ongoing research is centered on exploring the mechanisms behind emotional memory consolidation during sleep, with the goal of identifying potential new therapeutic targets for treating maladaptive processing of emotional memories and their early sleep-dependent consolidation.
www.science.org/doi/10.1126/science.adk3894

eppendorf & Science PRIZE FOR NEURO BIOLOGY

NEUROSCIENCE

To "feel" better, sleep on it!

Emotional memories are consolidated during REM sleep

By **Mattia Aime**[1,2]

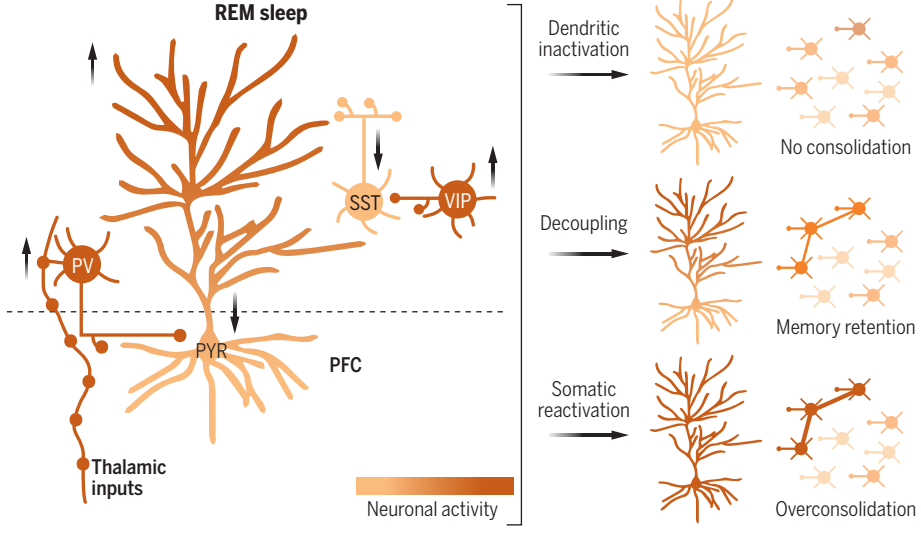
Emotions are a powerful driving force in our lives that influence our perceptions, behaviors, and overall well-being. As we navigate through the complexities of daily existence, we encounter a multitude of events that evoke strong positive and negative emotional responses. Yet how do we effectively manage this influx of emotional information without becoming overwhelmed? Interestingly, the answer to this question might lie in a seemingly unrelated phenomenon: sleep.

Sleep is an integral part of our daily routine, and, among the myriad functions attributed to this state, emotional processing stands out as a crucial aspect. Everyone has experienced at least once how a poor night's sleep can wreak havoc on our emotions. We have all faced the grumpiness, irritability, and even aggression that can come with sleep deprivation. Yet despite this common experience, the science behind how sleep and emotions are interwoven remains only partially understood.

Although evidence for the neural mechanisms associated with non–rapid eye movement (NREM) sleep (e.g., memory replay, synaptic homeostasis) has accumulated over the years, the function of rapid eye movement (REM, or paradoxical) sleep is less clear. In recent years, several studies have identified an important role of REM sleep in processing the information linked to emotional events encountered during the awake state (*1*–*3*). REM sleep is a state of high emotional content, characterized by vivid dreaming and heightened activity in brain regions associated with emotional processing, such as the amygdala and hippocampus (*4*, *5*). This heightened activity is thought to support the consolidation and integration of emotional experiences and aid in emotional memory formation. In addition to these brain regions, which are involved in emotional processing, the pre-

Somatodendritic decoupling optimizes emotional memories during REM sleep

During rapid eye movement (REM) sleep, the neuronal microcircuits in the prefrontal cortex (PFC) undergo a shift in excitation-inhibition balance, resulting in a somatodendritic decoupling of pyramidal (PYR) neurons. This mechanism is regulated by an REM-selective redistribution of activity of PV, SST, and VIP interneurons, which is partially driven by thalamic inputs. The somatodendritic decoupling is crucial for optimizing the processing of emotional memories. Optical inactivation of dendritic activity during REM sleep impairs the storage of emotional information, whereas somatic reactivation leads to memory overconsolidation.



PV, parvalbumin; SST, somatostatin; VIP, vasoactive intestinal peptide.

CREDITS: (PHOTO) COURTESY OF MATTIA AIME; (GRAPHIC) MATTIA AIME ADAPTED BY C. BICKEL/*SCIENCE*

frontal cortex has been recognized as a key player in regulating and modulating emotions (*6*, *7*). However, this brain structure is paradoxically quiescent during REM sleep, displaying diminished activity and reduced connectivity with other brain areas (*8*–*11*).

How does the prefrontal cortex process and store emotional information despite being silent during REM sleep? To decipher the enigma behind this paradox, my colleagues and I first dissected the microcircuit dynamics of the prefrontal cortex during REM sleep in head-restrained, naturally sleeping mice using simultaneous two-photon calcium imaging and polysomnography (with electroencephalography and electromyography to determine sleep states) (*12*). Remarkably, during REM sleep, the cell bodies (somas) of pyramidal neurons exhibited minimal activation, yet the dendritic arbors of these neurons displayed heightened activity, which resulted in a notable functional dissociation between these two subcellular compartments that we define as somatodendritic decoupling.

This mechanism is reminiscent of a simultaneous inhibition at the cell somas and disinhibition of their dendritic arborizations. To test this possibility, we recorded the main subpopulations of inhibitory neurons in the prefrontal cortex. Consistent with canonical circuit models, we identified a heightened activity of the parvalbumin (PV) interneurons, which are known to provide perisomatic inhibition of pyramidal neurons, during REM sleep (see the figure). Concurrently, we observed an REM-selective potentiation of inhibitory tone exerted by vasoactive intestinal peptide neurons onto the somatostatin interneurons. This process created a state of reduced inhibition, effectively disinhibiting the pyramidal dendrites and facilitating their activity (see the figure). Additionally, we found that long-range projections from the thalamus are predominantly responsible for promoting PV activity during REM sleep and consequently drive the somatic silencing of pyramidal neurons. To our knowledge, this is the first evidence that implicates thalamocortical circuits in REM sleep functions that go beyond their well-studied role in NREM sleep. Whether this mechanism can be generalized to the whole cortex, possibly through neuromodulation, awaits further investigation.

Does the somatodendritic decoupling play a role in the processing of emotional information during REM sleep? To explore this question, we used optogenetics to selectively block the somatodendritic decoupling during REM sleep after an auditory fear-conditioning discriminative learning task during which mice are trained to associate auditory cues with danger or safety stimuli (*12*). This perturbation of the decoupling was achieved either by optogenetic silencing of PV neurons during REM sleep and consequently reactivating the somatic compartment of pyramidal neurons or by suppressing the activity of VIP neurons and thus obstructing the dendritic disinhibition. REM-specific suppression of dendritic activity led to a loss of danger-versus-safety discrimination during learning and a lack of synaptic plasticity. However, optogenetic release of somatic inhibition led to a pyramidal neuronal activation that resulted in enhanced discrimination and synaptic potentiation. Hence, REM sleep–associated somatodendritic decoupling promotes opposing synaptic plasticity mechanisms that optimize emotional responses to future behavioral stressors (see the figure).

Notably, optically induced somatic reactivation of pyramidal neurons during REM sleep resulted in the overconsolidation of emotional memories (see the figure). Indeed, when mice were subsequently trained in a fear extinction task, they failed to adequately suppress the previously acquired fear memory. This inability to extinguish a fear memory is reminiscent of a pathological condition: posttraumatic stress disorder (PTSD). PTSD is the most common psychiatric disorder that arises after exposure to a traumatic event. Of note, among people who have experienced a traumatic event, those who suffer from major sleep problems are more likely to develop PTSD in the future. Indeed, one of the earliest signs of PTSD is sleep disturbance, which often includes nightmares, insomnia, and fragmented REM sleep (*13*, *14*). Thus, the interaction between REM-dependent stabilization of long-term emotional memories and the onset of PTSD is of great interest and relevance to the worldwide prevalence of this neuropsychiatric condition. With all this in mind, it is reasonable to hypothesize that a functional somatodendritic decoupling is compromised in PTSD patients. These findings open new perspectives for identifying therapeutic targets to treat maladaptive processing of traumatic memories, such as PTSD and their early (i.e., REM) sleep–dependent consolidation.

Changes in REM sleep quantity or quality are frequently symptomatic of affective disorders, including PTSD. Long-term improvement of sleep leads to decreased anxiety, stress, and depression, as well as increased well-being. Emotions have a powerful impact in our lives; so to "feel" better, let's sleep on it! ■

[1]Zentrum für Experimentelle Neurologie, Department of Neurology, Inselspital University Hospital Bern, Bern, Switzerland. [2]Department of Biomedical Research, University of Bern, Bern, Switzerland. Email: mattia.aime@unibe.ch

REFERENCES AND NOTES

1. I. C. Hutchison, S. Rathore, *Front. Psychol.* **6**, 1439 (2015).
2. M. Nishida, J. Pearsall, R. L. Buckner, M. P. Walker, *Cereb. Cortex* **19**, 1158 (2009).
3. S. Groch, I. Wilhelm, S. Diekelmann, J. Born, *Neurobiol. Learn. Mem.* **99**, 1 (2013).
4. G. Girardeau, I. Inema, G. Buzsáki, *Nat. Neurosci.* **20**, 1634 (2017).
5. R. Boyce, S. D. Glasgow, S. Williams, A. Adamantidis, *Science* **352**, 812 (2016).
6. M. Aime *et al.*, *eLife* **9**, e62594 (2020).
7. C. S. W. Lai, T. F. Franke, W. B. Gan, *Nature* **483**, 87 (2012).
8. E. A. Nofzinger, M. A. Mintun, M. Wiseman, D. J. Kupfer, R. Y. Moore, *Brain Res.* **770**, 192 (1997).
9. P. Maquet *et al.*, *Nat. Neurosci.* **3**, 831 (2000).
10. R. Maciel *et al.*, *Biochem. Pharmacol.* **191**, 114514 (2021).
11. A. Muzur, E. F. Pace-Schott, J. A. Hobson, *Trends Cogn. Sci.* **6**, 475 (2002).
12. M. Aime *et al.*, *Science* **376**, 724 (2022).
13. A. L. A. Murkar, J. De Koninck, *Sleep Med. Rev.* **41**, 173 (2018).
14. E. F. Pace-Schott, A. Germain, M. R. Milad, *Biol. Mood Anxiety Disord.* **5**, 3 (2015).

10.1126/science.adk3894

PRIZE ESSAY

FINALIST

Michael A. Skinnider

Michael Skinnider received his undergraduate degree from McMaster University and an MD/PhD from the University of British Columbia, during which time he was also a visiting PhD student at the École Polytechnique Fédérale de Lausanne. He then started his laboratory in the Lewis-Sigler Institute for Integrative Genomics and the Ludwig Institute for Cancer Research at Princeton University. His research focuses on the application of machine learning to problems in biology, chemistry, and medicine.

www.science.org/doi/10.1126/science.adk3912

eppendorf & Science PRIZE FOR NEURO BIOLOGY

NEUROSCIENCE

From single cells to neural circuits

Neural circuits are mapped in high throughput with single-cell genomics

By **Michael A. Skinnider**[1,2]

The human brain is composed of billions of neurons, wired together into neural circuits by trillions of synapses. Deciphering the organization of these neural circuits, and how they allow for the processing of information and the execution of complex behaviors, is a fundamental goal of neuroscience. Historically, however, dissecting neural circuits has been a low-throughput endeavor. Neuroscientists had to carefully synthesize existing knowledge to formulate hypotheses about the roles of specific neurons and then deploy painstaking experiments to test those hypotheses (*1*).

Over the past two decades, a series of technological advances has markedly accelerated the pace at which neural circuits can be dissected. Among the most exciting of these advances has been the advent of single-cell transcriptomics. Single-nucleus RNA sequencing (snRNA-seq) can measure the expression of thousands of genes across tens of thousands of neurons in a single experiment. This technology has opened an unprecedented window into neuronal diversity in the brain and has revealed hundreds of molecularly distinct neuronal subtypes (*2*, *3*). But snRNA-seq produces fundamentally static maps of gene expression. Linking neuronal subpopulations defined by single-cell transcriptomics to specific neural circuits—e.g., those activated by stimuli such as hunger or fear—has remained challenging.

I came face-to-face with this challenge in my efforts to understand the neural circuits engaged by epidural electrical stimulation (EES). EES, which involves delivering bursts of electricity to the lumbar spinal cord using an implantable device, has been shown to instantly restore walking in patients paralyzed by spinal cord injuries (*4*–*7*). Unexpectedly, EES was also found to mediate longer-term recovery, to the point that several patients eventually recovered the ability to walk even after the EES device was disabled. However, the neural circuits underlying this remarkable recovery remained enigmatic.

My colleagues at the École Polytechnique Fédérale de Lausanne and I sought to identify the neurons that could restore walking after paralysis (*8*). In the absence of a specific hypothesis about their identity, we surmised that we would need to simultaneously measure the molecular responses to EES of every neuronal subpopulation in the spinal cord. Single-cell genomics provided the ideal platform to achieve this goal. Using snRNA-seq, we measured gene expression in 20,990 spinal cord neurons from mice subjected to various modes of EES. However, leveraging these data to identify the specific neuronal subpopulations activated by EES proved more challenging than anticipated. Classical markers of neuronal activation, such as the immediate early gene c-*fos*, were expressed so lowly that they were almost indistinguishable from noise.

Clearly, a new approach was needed. I reasoned that neuronal responses to a given stimulus are not limited to the activation of immediate early genes and sought instead to develop a method that could take the entire neuronal transcriptome into account. My key hypothesis was that neurons that undergo a profound transcriptional response to EES should become separated from their unstimulated counterparts within the multidimensional space of gene expression. Aware of the limitations of widely used measures of separability in single-cell data (*9*), I saw an opportunity to overcome these limitations using machine learning. I posited that a machine-learning model could be trained to predict whether a given cell came from the spinal cord of a mouse that received EES or from that of a control mouse. Then, knowing which mouse each cell came from, we could calculate the accuracy of these predictions. For neurons that were strongly activated by EES, I thought that the model should readily learn to predict whether a particular cell came from a

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [2]Ludwig Institute for Cancer Research, Princeton University, Princeton, NJ, USA. Email: skinnider@princeton.edu

PHOTO: SAMEER KHAN

stimulated or unstimulated spinal cord. By contrast, for neurons not engaged by the therapy, the model's guesses should be no better than random chance.

I implemented these ideas in a method that I named Augur (*10*). Initially, I tested Augur with simulated single-cell data and was excited to find that Augur correctly prioritized the activated cell types in every simulation. Moreover, in published single-cell data, I found that Augur was able to correctly recapitulate well-studied neural circuits. In the visual cortex, for example, Augur accurately identified the neural circuits activated in response to light.

The ultimate test of Augur, however, would be its ability to help identify the neurons activated by EES. Remarkably, when Augur was applied to our snRNA-seq dataset, it prioritized the same subpopulation of visual system homeobox 2 (*Vsx2*)–expressing ventral excitatory interneurons in response to every mode of EES. This observation was further strengthened by spatial transcriptomics data from the same mice. I extended the logic underlying Augur into the setting of spatial transcriptomics and captured this new spatial logic in a second method that I named Magellan. I found that Magellan circumscribed the response to EES within the intermediate laminas of the spinal cord—exactly where the interneurons that Augur prioritized reside.

To elucidate the functional and anatomical properties of the prioritized neurons, we leveraged whole brain–spinal cord imaging of injured tissues, population-specific neuronal tracing, neuron-specific cellular recordings, cell ablation, chemogenetics, and ultimately optogenetics. These experiments unequivocally demonstrated that a single neuron subpopulation is both necessary and sufficient for walking after paralysis. We had indeed found the neurons that restore walking after paralysis.

In our hands, Augur and Magellan have been transformative tools to dissect neural circuits in the spinal cord and understand how they can be therapeutically targeted after spinal cord injuries. For example, we have now applied Augur and Magellan to discover a subpopulation of neurons that leads to loss of blood pressure control after spinal cord injury and another subpopulation that can be regrown through an anatomically complete injury to reestablish communication between the brain and spinal cord (*11*). Augur and Magellan have also helped in the dissection of cellular responses to perturbations across the central nervous system more broadly, including those of glial, vascular, and peripheral immune cells (*12*). Together, these methods provide a framework to tackle a fundamental challenge in neuroscience using unbiased single-cell technologies, and my hope is that they will accelerate the pace at which the neural circuits underlying any arbitrary behavior can be mapped. ■

REFERENCES AND NOTES

1. L. Luo, E. M. Callaway, K. Svoboda, *Neuron* **57**, 634 (2008).
2. J.-F. Poulin, B. Tasic, J. Hjerling-Leffler, J. M. Trimarchi, R. Awatramani, *Nat. Neurosci.* **19**, 1131 (2016).
3. BRAIN Initiative Cell Census Network (BICCN), *Nature* **598**, 86 (2021).
4. S. Harkema *et al.*, *Lancet* **377**, 1938 (2011).
5. C. A. Angeli *et al.*, *N. Engl. J. Med.* **379**, 1244 (2018).
6. M. L. Gill *et al.*, *Nat. Med.* **24**, 1677 (2018).
7. F. B. Wagner *et al.*, *Nature* **563**, 65 (2018).
8. C. Kathe *et al.*, *Nature* **611**, 540 (2022).
9. M. A. Skinnider, J. W. Squair, L. J. Foster, *Nat. Methods* **16**, 381 (2019).
10. M. A. Skinnider *et al.*, *Nat. Biotechnol.* **39**, 30 (2021).
11. J. W. Squair *et al.*, *Science* **381**, 1338 (2023).
12. M. A. Skinnider *et al.*, bioRxiv 10.1101/2023.06.23.544348 (2023).

10.1126/science.adk3912

RESEARCH

IN *SCIENCE* JOURNALS

Edited by
Michael Funk

GUT MICROBES

Good competition in the gut

Those who must use antibiotic prophylaxis to prevent opportunistic infections are at high risk for intestinal colonization and infection by multidrug-resistant organisms (MDROs). Woodworth *et al.* conducted a randomized controlled trial to compare the safety, efficacy, and strain dynamics of MDRO eradication after bowel preparation plus fecal microbiota transfer (FMT) versus bowel preparation alone in 11 renal transplantation recipients. FMT resulted in faster MDRO decolonization and protected study participants from recurrent infection. In some participants, extended spectrum beta lactamase (ESBL)–producing strains were replaced by non-ESBL strains, suggesting that strain competition rather than eradication may occur after FMT. —MLN

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abo2750

Gut microbes, pictured here in a colored scanning electron micrograph, can be replaced by fecal microbiota transfer, including strains that are drug resistant.

PHOTO: MARTIN OEGGERLI

MESOSCOPIC PHYSICS

Probing a chiral edge

One-dimensional systems of electrons behave in an unusual way that is described by the so-called Luttinger liquid theory. One way to study a one-dimensional system is to focus on the boundary of a two-dimensional fractional quantum Hall (FQH) state, which is home to chiral edge modes. The edge modes form chiral Luttinger liquids, which can be studied by tunneling electrons into them. To that end, Cohen *et al.* created a constriction known as a quantum point contact between an FQH state with 1/3 filling and an integer quantum Hall state with a filling of 1 in graphene. The conductance across the quantum point contact exhibited scaling laws characteristic of a chiral Luttinger liquid. —JS

Science, adf9728, this issue p. 542

GENOMICS

Arabidopsis nucleolus organizer regions

Nucleolus organizer regions contain ribosomal DNA (rDNA) repeats encoding the 18*S*, 25*S* to 28*S*, and 5.8*S* ribosomal RNAs (rRNAs) essential for protein synthesis. Fultz *et al.* used long-read DNA sequencing to provide a detailed structure of the two nucleolus organizer regions of *Arabidopsis thaliana*, NOR2 and NOR4, to study their expression and DNA methylation in flow-sorted nucleoli. The authors discovered many rRNA gene subtypes and found that most active genes are in central *NOR4*, whereas *NOR2* is mostly silent. This work provides a molecular understanding of nucleolar dominance, a process in which NORs are differentially expressed during development and NORs inherited from one progenitor are more active. Correlation of rDNA homogeneity with high expression implicates transcription in rDNA evolution. —DD

Sci. Adv. (2023) 10.1126/sciadv.adj4509

ACTIVE GALAXIES

A dense gas inflow feeding a black hole

Most galaxies contain a supermassive black hole at their center. When gas falls toward the black hole, it heats up and emits radiation and can be observed

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

GENE EXPRESSION

Sex, development, and gene expression

Because of differences in gene dosage compensation, hormone levels, and other factors, the expression of some genes is known to differ between sexes across sexually reproducing organisms. Rodríguez-Montes *et al*. looked at gene expression across organs through developmental stages for chicken, mouse, rat, rabbit, opossum, and humans to determine when differences seen in adulthood arise (see the Perspective by Sémon). Although the genes that are found on sex chromosomes or are involved in sex steroid pathways are often sex biased in at least one context, there are few genes with expression that is consistently sex biased across organs, development, or species. These results give insight into the rapid turnover of sex-biased gene expression over evolutionary time. —CNS

Science, adf1046, this issue p. 533; see also adk9990, p. 515

CHEMICAL BIOLOGY

Formaldehyde feedback

Exposure to concentrated exogenous formaldehyde is toxic, but organisms also produce considerable amounts endogenously, so cells must have mechanisms to sense and respond to formaldehyde stress. Pham *et al*. used activity-based profiling of mouse liver cells to identify potential privileged sites in the proteome that react with formaldehyde and might be involved in sensing overload. The authors identified a cysteine within an isoform of *S*-adenosylmethionine (SAM) synthase (MAT1A), that is preferentially modified and inhibited by formaldehyde exposure. The downstream effect in a mouse model of chronic formaldehyde stress was a reduction in SAM levels and a perturbation of specific histone methylation sites, which in turn led to increased transcription and protein levels of MAT1A. —MAF

Science, abp9201, this issue p. 532

IONTRONICS

Sending ionic signals to bullfrog hearts

Recognition and control of diverse bioionic signals, which would allow for the regulation of physiological processes in aqueous-phase biological media, remains one of the main challenges in the field of iontronics. State-of-the-art electronics and iontronics are still limited to electrons or single ions as signal carriers. Chen *et al*. report a heterogated biphasic gel iontronic device capable of biocompatible electronic-ionic signal processing and transmission, as demonstrated by regulating the cardiac electrical activity of bullfrog hearts. This approach is a promising step toward biocompatible signal processing and transmission in abiotic-biotic systems and has the potential to inspire interdisciplinary research at the intersection of electronics, ionics, chemistry, biology, and medicine. —YS

Science, adg0059, this issue p. 559

ELECTROCHEMISTRY

Conductivity control through occupancy

Lithium-metal-halides behave as superionic conductors and show stability at high operating voltages, thus making them suitable candidates for solid state batteries. Although several mechanisms for this behavior have been suggested, a detailed analysis of each one and its overall significance has been lacking. Yu *et al*. studied a family of hexagonal close-packed stacked trigonal halides with the general formula of Li_3MCl_6, where M is a metal ion such as yttrium (see the Perspective by Van der Ven). Using a combination of computational methods, the authors observed the importance of having partial occupancy of the metal in developing two critical structural characteristics: the existence of a percolating network and the maintenance of an appropriate interlayer distance. —MSL

Science, adg6591, this issue p. 573; see also adk8617, p. 513

NEUROSCIENCE

Mental time travel in virtual reality

The hippocampus holds a model of the environment that can be mentally traversed during recall or simulation. It is unknown whether animals can intentionally control their hippocampal activity according to their model of the world. By combining virtual reality and a real-time brain-machine interface, Lai *et al*. discovered that rats directly controlled their hippocampal neuronal firing in a goal-directed manner (see the Perspective by Coulter and Kemere). Rats first formed a hippocampal map of a virtual environment. Then, in brain-machine interface mode, they demonstrated the ability to activate representations from this map corresponding to specific remote locations, which then brought either them or an object to spatial goals. The rats could sustain a hippocampal representation of a remote location for tens of seconds, reminiscent of human imagination or mental time travel. —PRS

Science, adh5206, this issue p. 566; see also adl0806, p. 517

RESTORATION ECOLOGY

Herbivory constrains restoration success

Restoring vegetation to degraded areas, either through planting or by encouraging natural generation, is a prominent strategy for conservation and nature-based climate solutions. However, restoration efforts are not always successful and can take a long time to reach pristine conditions. Xu *et al*. performed a global meta-analysis to ascertain how herbivory affects restoration success in both terrestrial and aquatic systems (see the Perspective by Villar). They found that herbivory has negative effects on plant abundance and diversity at restoration sites, even more than in undisturbed ecosystems, and this effect was strongest at sites with actively planted vegetation. Their findings suggest that excluding herbivores or reintroducing predators may aid restoration efforts in many locations. —BEL

Science, add2814, this issue p. 589; see also adl0578, p. 516

WATER AVAILABILITY

Southern water

How has climate change affected the availability of land water over the recent past? Zhang *et al*. calculate global land water availability over the last two decades and found an overall decline that has been dominated by a negative trend in the Southern Hemisphere, whereas a mixture of positive and negative regional trends in the Northern Hemisphere have led to no significant change there (see the Perspective by Blöschl and Chaffe). El Niño is the most important climate mode affecting water availability in the Southern Hemisphere. —HJS

Science, adh0716, this issue p. 579; see also adk8164, p. 512

IMMUNOLOGY

Clonal eviction is key to tolerance

During their development, thymocytes with potentially autoreactive T cell receptors (TCRs) engage self-ligands presented by thymic epithelial cells with high affinity and are eliminated in a process called negative selection. Although TCR signaling is sustained during CD4 T cell selection

in the thymus, this is not the case for CD8 T cells, so autoreactive CD8 T cells must be managed differently. Badr *et al*. report in mice that immature autoreactive CD8 thymocytes are prematurely evicted from the thymus when strong TCR signaling downregulates the transcriptional repressor Gfi1 and induces the expression of S1P1. Thymocytes then travel to the periphery, where they develop into mature but self-tolerant clones. —STS

Science, adh4124, this issue p. 534

CANCER IMMUNOLOGY

Exhausted T cells turn off TEs

Transposable elements (TEs) are non–protein-coding DNA sequences capable of moving throughout the genome and composing nearly 50% of genomic DNA, but their function in T cells remains unclear. By analyzing bulk and single-cell transcriptomics, Bonté *et al*. explored the expression and regulation of TEs during the establishment of T cell exhaustion. In mouse tumor and chronic viral infection models, TEs belonging to the VL30 family were highly repressed in terminally exhausted, tumor-infiltrating T cells and were controlled by the transcription factor Fli1. In human tumors, TE expression patterns were associated with the state of T cell exhaustion and reprogrammed by anti–PD-1 immunotherapy. These findings demonstrate that TE expression is tightly regulated during the progression of T cell exhaustion, which could improve the accuracy of T cell gene signatures. —CO

Sci. Immunol. (2023) 10.1126/sciimmunol.adf8838

CELL BIOLOGY

Degrading endosomal redox signals

Fragments of extracellular matrix proteins stimulate chondrocytes to produce proteases, thus perpetuating cartilage degradation. Miao *et al*. found that the internalization of a fibronectin fragment and its integrin receptor induced primary human chondrocytes to increase their production of the protease MMP-13 (see the Focus by Goldring). This increase required the formation of redoxosomes, endosomes that are production sites for reactive oxygen species, and sustained intracellular integrin signaling. Cartilage from osteoarthritis patients showed increased formation of these redoxosomes. —AMV

Sci. Signal. (2023) 10.1126/scisignal.adf8299; see also 10.1126/scisignal.adj9760

as an active galactic nucleus. However, it is not known how gas is delivered from the wider galaxy to the inner region around the black hole. Izumi *et al*. combined submillimeter observations of several gas tracers in the center of the Circinus galaxy, a nearby active galactic nucleus. By comparing the kinematics of molecular, atomic, and ionized gas, they were able to identify an inflow of dense molecular gas in the central parsec, feeding the black hole. —KTS

Science, adf0569, this issue p. 554

THERMAL TRANSPORT

A quick thermal switch

Controlling thermal transport is challenging because of the difficulty in creating devices that work quickly and can be cycled many times. Li *et al*. developed a three-terminal thermal switch in which the heat flow is modulated by an electric field. The device works through careful control of chemical bonding and charge distribution. The resulting switching speeds are above 1 megahertz, with thermal conductance ratios over 1300%, and can be switched more than 10 million times. —BG

Science, abo4297, this issue p. 585

MPOX

Into the heart of defenselessness

In March 2022, an international epidemic of human Mpox was detected, showing that it was not solely a zoonotic infection. A hallmark of the approximately 88,000 cases that have been reported were TC→TT and GA→AA mutations in Mpox viruses, which were acquired at a surprisingly high evolutionary rate for a pox virus. Knowing that these types of mutation are a sign of activity by a host antiviral enzyme called APOBEC3, O'Toole *et al*. investigated whether the mutations reflected human-to-human transmission rather than repeated zoonotic spillover. Bayesian evolutionary analysis showed that Mpox virus recently diversified into several lineages in humans that display elevated numbers of mutations, signaling APOBEC exposure and sustained human-to-human transmission rather than zoonosis as the source of new cases. —CA

Science, adg8116, this issue p. 595

DEVELOPMENT

Flipping a switch in gonadal development

The development of ovaries or testes in mammals depends on the presence of specific supporting cells: Sertoli cells for the testes and pregranulosa cells for the ovaries. The gene *Sry*, located on the Y chromosome, plays a key role in promoting testicular development, but the driver(s) of ovarian development has remained unknown. Gregoire *et al*. have revealed the role of the tumor suppressor protein WT1 in ovarian development in mice. WT1 has two major isoforms that do or do not include three amino acids in a particular site in the protein. A change in the ratio of isoforms suppresses the expression of SRY, resulting in a shift toward ovarian rather than testicular development. —YN

Science, add8831, this issue p. 600

DIOXYGEN ACTIVATION

MOF mimic of an oxygenase

Redox-active metalloproteins are often adept at binding to and controlling the reactivity of oxygen. Hou *et al*. showed that a metal-organic framework can similarly activate dioxygen and perform hydroxylation of aliphatic substrates. The authors used a range of spectroscopic techniques to demonstrate that the mechanism proceeds through a high-spin iron(IV)-oxo species, similar to what is seen in some dioxygenase enzymes. This system can perform catalytic oxygenation of cyclohexane and should inspire future catalyst development for dioxygen activation and hydrocarbon oxidation. —MAF

Science, add7417, this issue p. 547

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

TOPOLOGICAL MATTER

Intertwining topology with interactions

Two of the most important topological states, quantum spin Hall and quantum anomalous Hall, have distinct characters and tend to appear in different materials. Recently, however, both states were observed in the same material, a $MoTe_2$/WSe_2 moiré bilayer, a result that could not be understood using standard noninteracting models. Inspired by these findings, Mai *et al.* applied analytical and numerical methods to study the interplay of topology and interactions in models of two-dimensional topological systems. Both a simpler, exactly solvable model and more realistic ones incorporating Hubbard interactions indicated the formation of topological states consistent with experiments. —JS

Nat. Commun. (2023) 10.1038/s41467-023-41465-6

FUNGI

Rocky radiation resisters

Organisms vary widely in their ability to withstand ionizing radiation, but for microbes growing on exposed rock surfaces, protection from the sun is essential. Aureli *et al.* collected data on fungal radiation resistance for 101 strains of black meristematic fungi from sites around the world, including the harsh, dry valleys of Antarctica. Consistent with

Financial incentives to reduce forest exploitation in China may still disadvantage the very poor.

ECOLOGICAL ECONOMICS

Improving policies for people in poverty

Programs that pay people for actions that produce ecosystem services, such as managing land to prevent soil erosion, have the potential to improve environmental conditions while also reducing poverty. Such programs must be carefully designed to help those most in need. Hu *et al.* used a cross-sectional dataset of 5225 households in Sichuan Province, China, to evaluate a program that pays farmers to give up land usage rights in forests to increase carbon storage. They found that whereas moderately poor households benefited from the program, extremely poor households with limited labor capacity suffered from the loss of forest resources without the opportunity for income from forest-tending work, suggesting a need for additional avenues to receive compensation. —BEL

J. Ecol. Econ. (2023)
10.1016/j.ecolecon.2023.107976

prior work, the authors found that adaptation to extreme environments in general may provide enhanced resistance to gamma radiation. Some fungi are also known to tolerate and possibly even derive energy from anthropogenic radiation sources, although all of these mechanisms require much more study. —MAF

Environ. Microbiol. (2023)
10.1111/1462-2920.16510

NATURAL PRODUCTS

Plant-based neuronal protection

Accumulation of proteins containing polyglutamine repeat (polyQ) in neurons leads to the formation of toxic aggregates and subsequent neurodegeneration in humans. PolyQ affecting human huntingtin (hHTT) leads to the development of Huntington's disease. Interestingly, plants express many polyQ-containing proteins, but no cell degeneration or plant pathologies have been observed. To investigate this discrepancy, Llamas *et al*. expressed polyQ-hHTT in plant cells and showed that the stromal processing peptidase (SPP) in chloroplasts interacted with the polyQ, preventing protein aggregation. Synthetic SPP expressed in human cells prevented polyQ-hHTT aggregation. Although very preliminary, this study suggests that we might be able to leverage plant cell biology to find therapies for mammalian neurodegeneration. —MMa

Nat. Aging (2023)
10.1038/s43587-023-00502-1

ARIDITY

Dry-field patterning

How will the global distribution of aridity change as anthropogenic activities continue to affect temperature and precipitation fields? Yin and Porporato found that the pattern of aridity is controlled primarily by the spatial variability of precipitation, that this pattern has not changed significantly in the past century, and that it is not expected to change even in a drier, warmer future. Knowing how aridity and water availability will change is key to understanding the Earth system in general and issues such as food security, ecosystem resilience, carbon storage, and public health in particular. —HJS

Geophys. Res. Lett. (2023)
10.1029/2023GL105228

CELL BIOLOGY

Going through a phase

The secretory pathway guides secretory and membrane proteins from their site of synthesis in the endoplasmic reticulum (ER) through the Golgi complex in transit to the cell exterior. Liquid-liquid phase separation (LLPS) can organize cellular compartmentalization, but the regulators of this process are largely elusive. Gallo *et al.* found that protein assemblies involved in packaging and transport of proteins from the ER membrane, so-called ER exit sites (ERES), are organized through LLPS. Counteracting activities of DYRK3 kinase and phosphatases establish a dynamic liquid-like environment within the ERES. These biophysical changes are required for efficient secretory protein trafficking from the ER to the Golgi. —SMH

Dev. Cell (2023)
10.1016/j.devcel.2023.08.005

CARDIOLOGY

Be still my beating heart

The heart continues to beat without pause throughout an animal's life. Even isolated cardiac cells and tissues maintain pulsations, which makes them useful models for investigations of heart biology. Unfortunately, pulsating cells are difficult to study because the constant motion can interfere with imaging-based methods of analysis. Cells can be paralyzed using a pharmacological compound such as blebbistatin to facilitate imaging, but inhibitors alter the cells' metabolism, which precludes the study of cardiac contraction in action. A computational approach from Woodhams *et al.* allows cardiac cells to keep on contracting and also allows simultaneous analysis by digitally removing motion artifact without altering the tissue itself. —YN

Proc. Natl. Acad. Sci. U.S.A. (2023)
10.1073/pnas.2212949120

RESEARCH ARTICLE SUMMARY

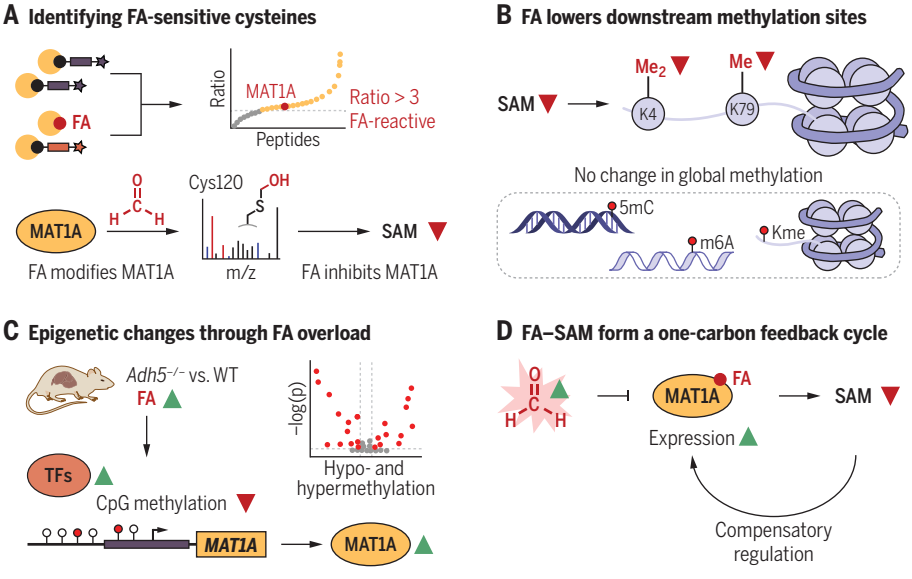
CHEMICAL BIOLOGY

Formaldehyde regulates *S*-adenosylmethionine biosynthesis and one-carbon metabolism

Vanha N. Pham†, Kevin J. Bruemmer†, Joel D. W. Toh, Eva J. Ge, Logan Tenney, Carl C. Ward, Felix A. Dingler, Christopher L. Millington, Carlos A. Garcia-Prieto, Mia C. Pulos-Holmes, Nicholas T. Ingolia, Lucas B. Pontel, Manel Esteller, Ketan J. Patel, Daniel K. Nomura, Christopher J. Chang*

INTRODUCTION: One-carbon metabolism manages cellular carbon pools by detoxifying highly reactive carbon species, such as aldehydes, and diverting their carbon toward the biosynthesis of useful products, including amino acids and nucleotides. Formaldehyde (FA) is a major one-carbon unit derived from exogenous environmental exposure and endogenous sources and is quickly scavenged in the cell through enzymatic oxidation to formate and carbon dioxide and/or metabolized through the folate cycle. *S*-adenosylmethionine (SAM) serves as the primary cellular methyl donor and harbors one-carbon units in a stable and accessible form. The ability to decipher the biochemical interplay between toxic reactive carbon species and stable physiological carbon units is essential for understanding fundamentals of one-carbon metabolism across all kingdoms of life. Especially important is understanding how aberrant carbon imbalances are connected to human diseases such as cancer, liver diseases, and asthma. Although the chronic exposure of toxic aldehydes is correlated to disease states, biological mechanisms of aldehyde signaling and their relation to carbon metabolism remain underexplored.

RATIONALE: Owing to its highly electrophilic nature, we hypothesized that FA could act as a one-carbon signal sensed by privileged cysteine sites across the proteome. FA reacts with cysteines on synthetic peptides, and we designed an unbiased, proteome-wide profiling study to systematically identify FA-sensitive cysteine residues. This work builds a biochemical framework for understanding global FA reactivity as a selective posttranslational modification of target proteins and downstream regulatory effects of such modifications.



FA-dependent inhibition of SAM biosynthesis. (**A**) ABPP identifies FA-sensitive cysteines such as MAT1A Cys120, in which FA modification inhibits *S*-adenosylmethionine synthase isoform type-1 (MAT1A) activity. (**B**) SAM deficiency lowers select histone methylation (Me) marks. (**C**) *Adh5*$^{-/-}$ mouse model of FA overload displays epigenetic alterations. (**D**) Inhibition of MAT1A activity by FA decreases SAM levels, which in turn promotes a one-carbon feedback cycle by compensatory increases in MAT1A expression. [Figure created using BioRender]

RESULTS: Activity-based protein profiling identified FA modification of privileged cysteine sites across the proteome, including several enzymes responsible for FA metabolism, one-carbon metabolism, and amino acid biosynthesis. We focused on biochemical characterization of a key Cys120 residue on the SAM-generating enzyme *S*-adenosylmethionine synthase isoform type-1 (MAT1A) that is proximal to the MAT1A active site. FA exposure resulted in inhibition of MAT1A activity in an isoform-specific manner, which led to decreased SAM production. Cellular models containing only the MAT1A isoform displayed a reciprocal decrease in SAM levels with increasing doses of FA exposure. Moreover, an *Adh5*$^{-/-}$ mouse model of chronic FA elevation also showed SAM deficiency accompanied by lower levels of methylation on select histone methyl sinks. The chronic FA model also resulted in a decrease in methylation of the *Mat1a* promoter region, resulting in increased MAT1A expression as a compensatory mechanism to maintain available carbon units. We deciphered a compensatory biochemical feedback cycle where FA-dependent SAM deficiency led to an increase in MAT1A expression through genetic and epigenetic mechanisms regulated by FA-dependent transcription factors and DNA promoter hypomethylation, respectively.

CONCLUSION: In contrast to the traditional view of FA as an indiscriminate electrophile and toxic metabolite, we show that FA is sensed by specific cysteine sites in the proteome to regulate one-carbon metabolism feedback cycles through SAM biosynthesis. FA reacts with a key cysteine residue on MAT1A to inhibit its activity, resulting in SAM depletion and downstream changes in histone and DNA methylation. Under normal homeostatic conditions, FA is quickly sequestered into the folate cycle for conservation of one-carbon units to maintain balanced SAM biosynthesis. In response to FA overload, reciprocal SAM depletion through isoform-specific MAT1A inhibition results in changes to cellular methylation potential, epigenetic dysregulation, and perturbations in one-carbon metabolism, which in turns leads to compensatory up-regulation of MAT1A expression. This work provides a starting point for further exploration of aldehydes as signaling agents and the nexus between one-carbon metabolism and one-carbon signaling. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: chrischang@berkeley.edu
†These authors contributed equally to this work.
Cite this article as V. N. Pham *et al.*, *Science* **382**, eabp9201 (2023). DOI: 10.1126/science.abp9201

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abp9201

RESEARCH ARTICLE

CHEMICAL BIOLOGY

Formaldehyde regulates *S*-adenosylmethionine biosynthesis and one-carbon metabolism

Vanha N. Pham[1]†, Kevin J. Bruemmer[1]†, Joel D. W. Toh[1], Eva J. Ge[1]‡, Logan Tenney[1], Carl C. Ward[2]§, Felix A. Dingler[5], Christopher L. Millington[5], Carlos A. Garcia-Prieto[6,7], Mia C. Pulos-Holmes[2], Nicholas T. Ingolia[2], Lucas B. Pontel[6,8], Manel Esteller[6,9,10,11], Ketan J. Patel[5], Daniel K. Nomura[1,2,3,4], Christopher J. Chang[1,2]*

One-carbon metabolism is an essential branch of cellular metabolism that intersects with epigenetic regulation. In this work, we show how formaldehyde (FA), a one-carbon unit derived from both endogenous sources and environmental exposure, regulates one-carbon metabolism by inhibiting the biosynthesis of *S*-adenosylmethionine (SAM), the major methyl donor in cells. FA reacts with privileged, hyperreactive cysteine sites in the proteome, including Cys120 in S-adenosylmethionine synthase isoform type-1 (MAT1A). FA exposure inhibited MAT1A activity and decreased SAM production with MAT-isoform specificity. A genetic mouse model of chronic FA overload showed a decrease n SAM and in methylation on selected histones and genes. Epigenetic and transcriptional regulation of *Mat1a* and related genes function as compensatory mechanisms for FA-dependent SAM depletion, revealing a biochemical feedback cycle between FA and SAM one-carbon units.

The one-carbon cycle is central to cell function, in which activation of one-carbon units is used for nucleotide and amino acid biosynthesis and epigenetic regulation (*1*, *2*). The one-carbon unit *S*-adenosylmethionine (SAM) is a ubiquitous cosubstrate and key methyl donor for transmethylation reactions that underlie one-carbon metabolic (*3*) and epigenetic processes (*4*), including modulating mTORC1 signaling for methionine sensing and cellular growth (*5*). Misregulation of SAM biosynthesis and metabolism is observed in diseases such as nonalcoholic fatty liver disease (*6*) and cancers (*1*, *3*, *7*).

Another one-carbon unit of emerging importance is formaldehyde (FA) (*8*, *9*). In addition to the role of FA as an environmental carcinogen, FA is endogenously produced through biological pathways, including enzymatic serine cleavage (*3*), oxidative folate degradation (*10*), and oxidative demethylation (*11*, *12*). Previously, activity-based sensing probes have shown the substantial generation of endogenous FA (*13*–*17*) with high relative abundance compared with other reactive carbon species such as aldehydes (*18*), in which the FA basal level in blood is 50 μM and can exceed 500 μM in disease states (*19*, *20*). FA is detoxified by a two-tier mechanism involving the enzyme alcohol dehydrogenase 5 (ADH5) for FA removal and the Fanconi anemia pathway for FA-induced DNA crosslinking repair, where loss of FA scavenging promotes carcinogenesis (*21*, *22*). In this work, we report the discovery of a biochemical feedback cycle between FA and SAM one-carbon units. Specifically, we identified FA dose-dependent inhibition of SAM production through targeting *S*-adenosylmethionine synthase isoform type-1 (MAT1A), the terminal enzyme in SAM biosynthesis, at a privileged, isoform-specific Cys120, to decrease the overall methylation potential of the cell. Contrary to the conventional view of FA as a promiscuous electrophile, this work reveals that this one-carbon signal can participate in site-specific post-translational modifications to enable broader biochemical regulation of central metabolic functions in the cell.

[1]Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720, USA. [2]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA. [3]Department of Nutritional Sciences and Toxicology, University of California, Berkeley, Berkeley, CA 94720, USA. [4]Innovative Genomics Institute, Berkeley, CA 94704, USA. [5]MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [6]Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Catalonia, Spain. [7]Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain. [8]Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA), CONICET-Partner Institute of the Max Planck Society, Buenos Aires, Argentina. [9]Centro de Investigacion Biomedica en Red Cancer (CIBERONC), Calle Monforte de Lemos, Madrid, Spain. [10]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluis Companys, Barcelona, Spain. [11]Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona, Feixa Llarga, l'Hospitalet de Llobregat, Spain.

*Corresponding author. Email: chrischang@berkeley.edu

†These authors contributed equally to this work.

‡Current affiliation: Department of Chemistry, University of North Carolina at Charlotte, Charlotte, NC, USA.

§Current affiliation: Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA.

Results

Activity-based protein profiling chemoproteomics reveals privileged, FA-sensitive cysteine targets in one-carbon metabolism

Motivated by in vitro observations that FA reacts with cysteine residues on peptides to form hemithioacetal and thiazolidine adducts (*23*, *24*), and related reactive electrophilic species emerging as key players in redox signaling and stress pathways (*18*), we sought to identify FA-sensitive targets across the proteome in an unbiased, high-throughput manner. We applied the isotopic tandem orthogonal proteolysis–activity-based protein profiling (isoTOP-ABPP) platform (*25*) to mouse liver lysates exposed to 500 μM of FA—close to disease-level concentrations (*20*)—or vehicle control. By using the cysteine-reactive iodoacetamide (IA)–alkyne probe followed by copper-catalyzed azide-alkyne cycloaddition (CuAAC) to enable attachment of either light (vehicle-treated) or heavy (FA-treated) isotopically labeled Tobacco Etch Virus (TEV) protease–cleavable biotin enrichment tags (fig. S1) (*26*), liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis afforded quantified ratios of light- versus heavy-modified peptides. Sites with higher incorporation of IA-alkyne on vehicle-treated light versus FA-treated heavy labels indicated higher FA reactivity at specific cysteine residues of identified protein targets (Fig. 1A).

Proteomic ratio analyses produced a list of 576 probe-modified peptides with an isotopic ratio greater than 3, equivalent to a $\log_2$ ratio greater than 1.58 (Fig. 1B and table S1). In contrast to the conventional view of FA as an indiscriminate electrophile, the data revealed privileged FA-sensitive cysteine sites, similar to other reactive species such as hydrogen peroxide, nitric oxide, and methylglyoxal (*27*, *28*), and provides foundational information on FA-cysteine reactivity across the proteome in a systematic manner. Notably, modified cysteines were found enriched in protein families spanning FA-dependent pathways (e.g., carbon metabolism and glutathione) and one-carbon metabolism (e.g., one-carbon amino acids and folate) rather than demonstrating proteome-wide, indiscriminate reactivity (Fig. 1C). Included are targets such as a key enzyme in FA detoxification, alcohol dehydrogenase 5 (class III) (ADH5) (*21*); a known regulator of methyl unit balance through the biosynthesis of sarcosine from glycine, glycine-*N*-methyltransferase (GNMT) (*29*); and serine hydroxymethyltransferases 1 and 2 (SHMT1 and SHMT2) that regulate serine and glycine pools in the folate cycle. Additional targets identified in amino acid metabolism, methyl transfer and transsulfuration chemistry, and carbohydrate metabolism point to the specificity with which FA engages enzymes known to regulate methyl units and the one-

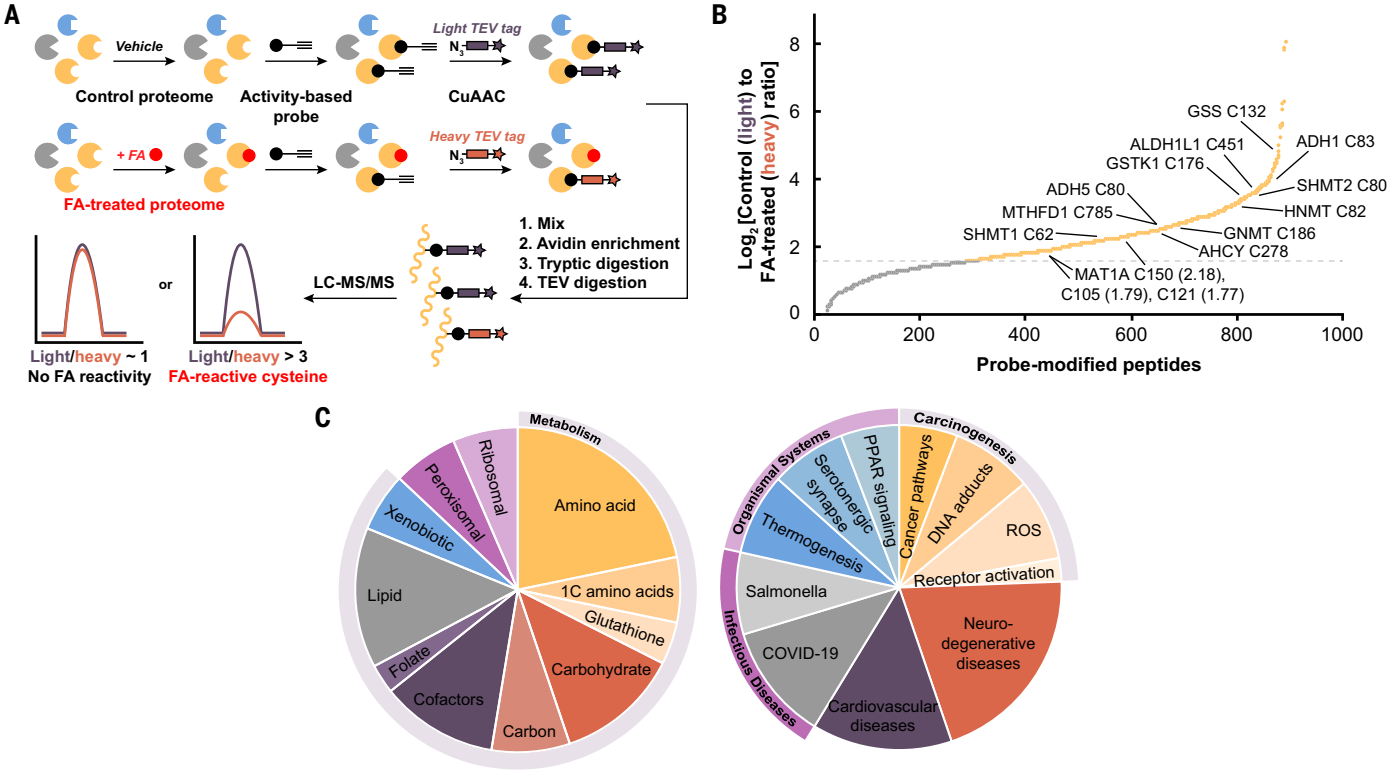


Fig. 1. isoTOP-ABPP identifies privileged, FA-sensitive cysteine sites across the proteome. (**A**) Workflow for isoTOP-ABPP analysis of FA-sensitive cysteine sites applied to whole proteomes. Mouse liver lysate was treated with vehicle or FA, followed by IA-alkyne cysteine activity-based probe, labeling of isotopic tags with CuAAC, downstream digestion, and subsequent analysis of peptide fragments with LC-MS/MS. (**B**) Waterfall plot of light to heavy ratios of FA-sensitive cysteine sites from isoTOP-ABPP (yellow), revealing a pattern of privileged targets for this one-carbon unit. Important enzymes involved in one-carbon and FA metabolism are labeled with the cysteine residue and MAT1A ratios in parentheses. The gray dotted line represents $\log_2$ ratio of 1.58, equivalent to a ratio of 3. Targets were filtered for appearing in at least two technical replicates. (**C**) Pie chart summarizing the KEGG pathways most abundant with enzymes with ratios >3 for FA-sensitive cysteines. Pathways were separated into two pie charts by metabolism and cellular processes (left; 308 proteins) and organismal systems and diseases (right; 172 proteins).

carbon cycle in a variety of cellular contexts. Of particular interest is MAT1A, a key enzyme in SAM biosynthesis that regulates the flux of transmethylation reactions. Out of the 10 total cysteine residues in MAT1A, isoTOP-ABPP analysis identified three specific ones that were hyperreactive toward FA: mouse MAT1A Cys105, Cys121, and Cys150, with ratios of 3.45, 3.40, and 4.54, respectively.

FA makes covalent adducts with MAT1A

We next sought to directly identify sites of FA modification on human MAT1A through mass spectrometry. We treated purified MAT1A protein with excess FA (5 mM) for 4 hours at 37°C and then digested overnight into peptides. When added in large excess, FA forms modifications on arginine, lysine, tryptophan, histidine, and cysteine in simple in vitro contexts of model peptides and free amino acids (*23*, *24*). However, LC-MS/MS quantification of MAT1A modified peptides and an unbiased mass search revealed limited FA reactivity with nucleophilic amino acid residues beyond cysteine (Fig. 2A). Only the three hyperreactive cysteine residues identified by isoTOP-ABPP (human Cys104, Cys120, and Cys149, analogous to mouse Cys105, Cys121, and Cys150) were modified by excess FA in vitro, along with a fourth solvent-accessible cysteine residue (human Cys376), indicating high selectivity of FA toward these privileged cysteine sites compared with other nucleophilic amino acids (Fig. 2, B to E). Modification of Cys104 indicated a mass-to-charge ratio (m/z) change of +12 Da, corresponding to nucleophilic addition to FA by the cysteine thiol to form a hemithioacetal, and cyclization to form a thiazolidine (fig. S2) (*24*). By contrast, Cys120 (fig. S3), Cys149 (fig. S4), and Cys376 (fig. S5) showed an m/z change of +30 Da, corresponding to a hemithioacetal modification (*24*). The labeling pattern observed in vitro with purified protein further establishes that FA is not a global, indiscriminate electrophile and validates the isoTOP-ABPP chemoproteomics method in identifying privileged cysteine sites that are selective to FA.

To determine if modification was selective for FA over other aldehydes, we performed the modification identification on purified MAT1A with excess acetaldehyde (5 mM) (*30*). Quantification of the same peptide coverage as with FA revealed no discernable modifications of any amino acid residue with acetaldehyde. The lack of acetaldehyde modification is supported by data that shows that FA reacts the fastest and forms the most stable product with cysteine compared with other carbonyl compounds (*24*), which supports the specificity of FA as a privileged electrophile for protein posttranslational modifications.

Attempts to characterize covalent FA-MAT1A modifications by protein crystallography have not yet been successful, in part because of the susceptibility of FA adducts to long-term hydrolysis (*24*). We were able to obtain a crystal structure of MAT1A at 2.00-Å resolution, which revealed SAM bound in the active site along with additional cofactors not mapped in previous structures (*31*), including imidotriphosphate (PPNP), Mg^{2+} ions, and K^{+} ions. Cys120 lies closest in proximity to the MAT1A active site, presaging this residue as the functional site of FA reactivity (Fig. 2F, fig. S6, and table S2).

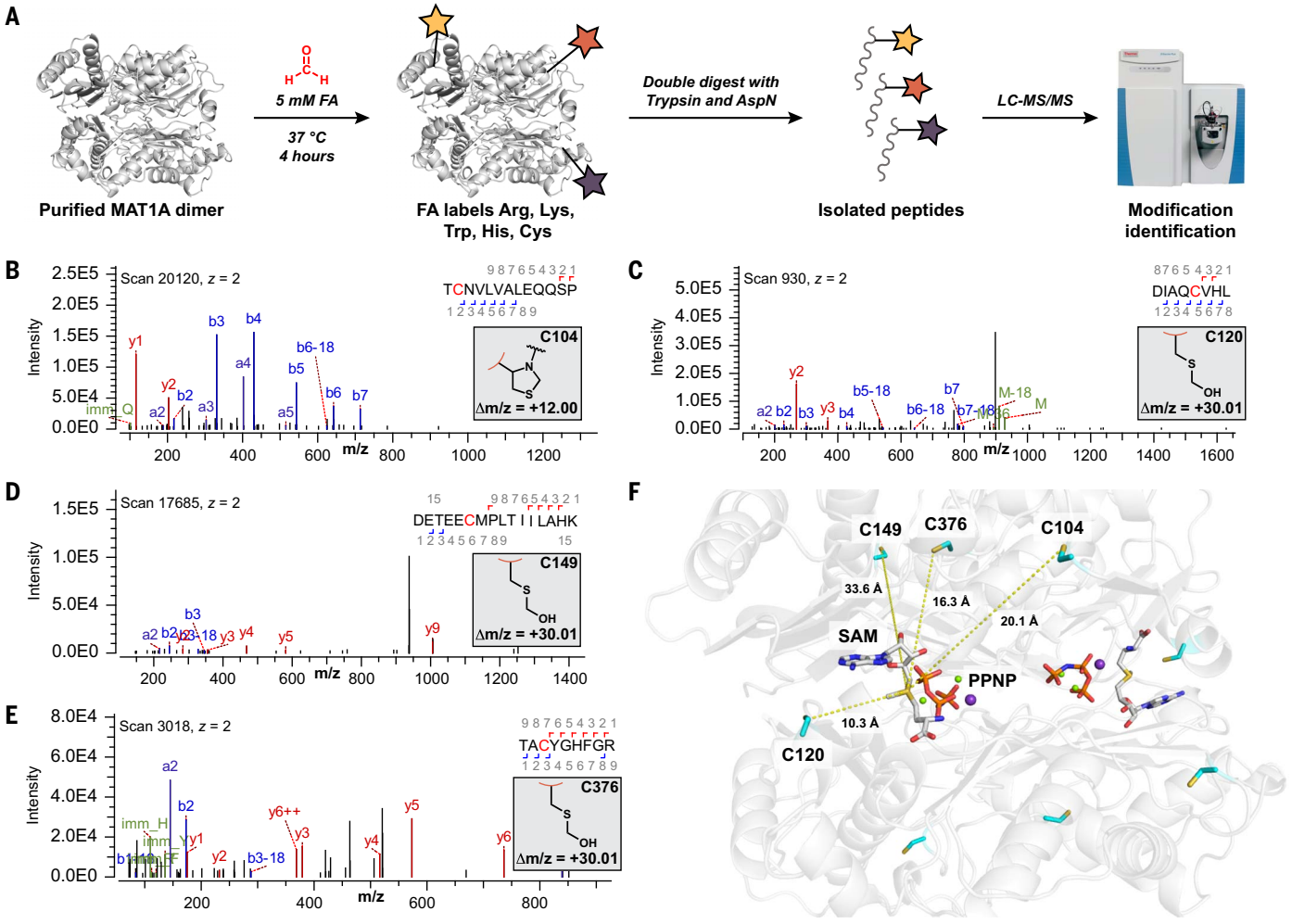


Fig. 2. FA forms site-specific covalent cysteine modifications on MAT1A. (**A**) Workflow for the identification of FA-dependent covalent modifications on purified MAT1A protein in vitro. AspN, endoproteinase that cleaves N-terminal to Asp residues. (**B** to **E**) Representative MS/MS spectra of FA modification of (B) Cys104, (C) Cys120, (D) Cys149, and (E) Cys376 upon fragmentation and sequencing to generate *b* (blue) and *y* (red) ions to indicate a (B) thiazolidine (Δm/z = +12.00) and (C to E) hemithioacetal (Δm/z = +30.01) modifications, respectively. (**F**) Crystal structure showing the position of all four identified cysteine residues relative to the SAM-bound active site of MAT1A. The angstrom distance between each cysteine residue and SAM are noted. SAM and PPNP, substitute for ATP, are labeled.

Additionally, previous studies have reported Cys120 as a target of hydrogen peroxide and nitric oxide inhibition of rat MAT1A (*32*–*34*).

FA inhibits MAT1A but not MAT2A activity in vitro at Cys120

To characterize the biochemical effects of FA-modified cysteine sites on MAT1A and establish the role of FA in SAM biosynthesis, we investigated the effects of FA on the activity of both MAT1A and MAT2A isoforms. Notably, an isoform switch from MAT1A to MAT2A occurs during hepatocellular carcinogenesis, resulting in intracellular SAM depletion in tumor environments (*35*, *36*). The cancer-dependent isoform switch, coupled with the high product inhibition by SAM for MAT2A compared with MAT1A, makes MAT2A a promising therapeutic target for blocking tumor growth by increasing the bioavailability of SAM (*37*).

We purified catalytically active forms of MAT1A (tetramer) and MAT2A (dimer). SAM biosynthesis occurs through three genes (*MAT1A*, *MAT2A*, and *MAT2B*) encoding MAT enzymes, which catalyze the nucleophilic addition of L-methionine to adenosine triphosphate (ATP) (*38*). *MAT1A* encodes for expression of tetramer (MAT1A) and homodimer (MATIII), which are primarily expressed in healthy liver tissue, whereas *MAT2A* and *MAT2B* encode for expression of MAT2A and MATβ regulatory subunit, respectively, which are expressed in all other cell types, fetal liver tissue, and liver cancer (*38*). Although MAT1A and MAT2A share high sequence homology (84%) (*39*), they differ in substrate affinity and product inhibition, which affects their levels of SAM production (supplementary text) (*40*, *41*).

To determine a physiologically relevant range of FA concentrations, we measured basal FA levels in HepG2 cells using a ratiometric fluorescent probe previously developed in our laboratory, RFAP-1 (*16*). Standard addition from exogenously added FA allowed us to obtain a cellular FA concentration of 49 ± 10 μM in HepG2 (Fig. 3A) comparable to reported measurements (*19*, *20*, *22*, *42*). To probe the effects of FA on MAT1A activity, we performed kinetic enzyme assays to monitor SAM production in vitro using triple quadrupole (QqQ) mass spectrometry (*37*). Enzymatic activity of MAT1A decreased 44% in response to 100 μM FA and 60% in response to 500 μM FA (Fig. 3B). By contrast, MAT2A retained full catalytic activity in the presence of even large excess of FA at 100 and 500 μM (Fig. 3B). FA thus selectively inhibits MAT1A activity in an isoform-specific manner.

Cys104 and Cys149 are common to both MAT isoforms, whereas Cys120 and Cys376 are present only in MAT1A. To identify which of the cysteine residues imparts sensitivity to FA-dependent catalytic activity inhibition, we

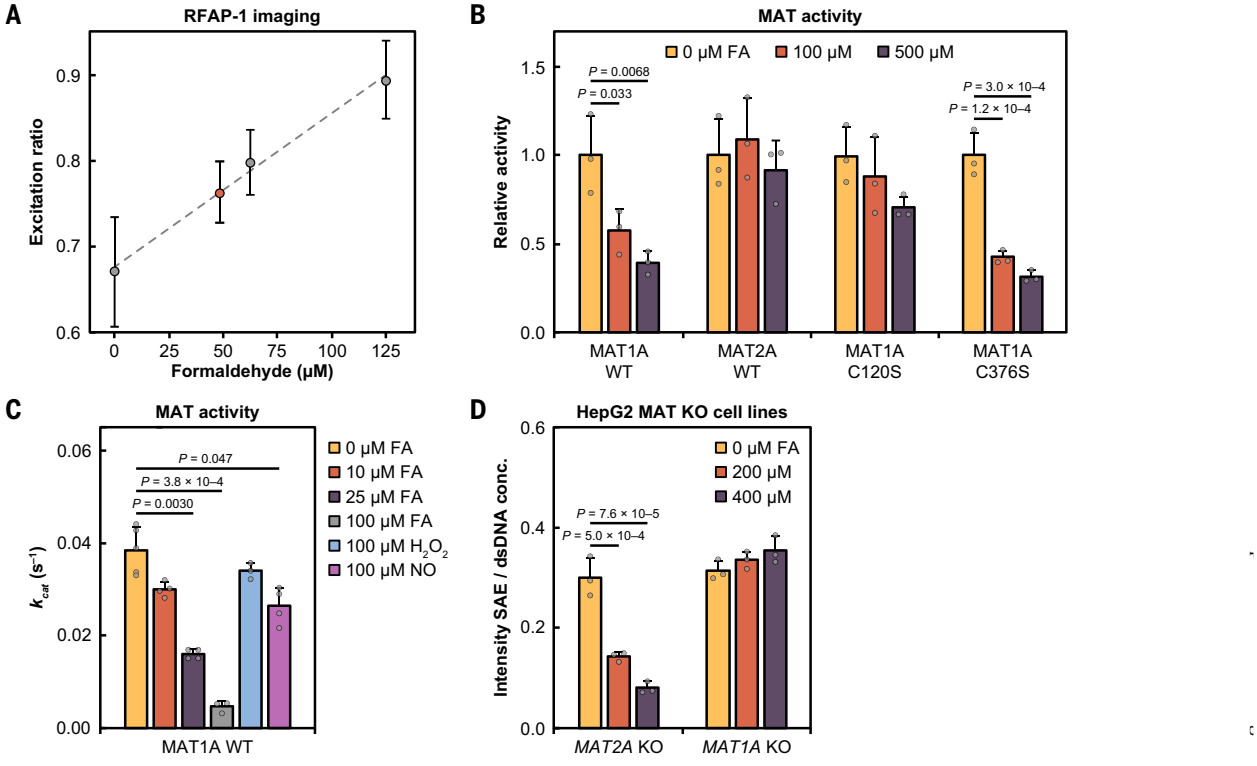


Fig. 3. FA inhibits SAM biosynthesis of MAT1A at Cys120 on purified protein in a dose- and isoform-dependent manner, lowering levels of SAM in cells. (**A**) Standard addition curve of FA in HepG2 by RFAP-1 fluorescent probe. Error bars represent the SD (n = 4) of technical cell replicates for each concentration. (**B**) Relative activity of MAT1A WT, MAT2A WT, MAT1A C120S, and MAT1A C376S enzymes in response to 0 (yellow), 100 (orange), or 500 μM (purple) FA. Error bars represent the SD (n = 3) of independent protein aliquots. (**C**) k_{cat} of MAT1A WT untreated or treated with 10 μM, 25 μM, and 100 μM of FA, and 100 μM of reactive species (H_2O_2 or NO). Michaelis-Menten kinetic analysis and K_M values are reported in table S3. Error bars represent SD (n = 3 or 4) of independent protein aliquots. (**D**) Measurements of SAE levels in CRISPR-generated HepG2 cell models showing that MAT1A-positive *MAT2A* KO cells are sensitive to FA and lead to SAE depletion, whereas SAE levels in MAT2A-positive *MAT1A* KO cells are unaffected by FA exposure. Error bars represent SD (n = 3) of technical cell replicates. dsDNA, double-stranded DNA. Statistical significance (B to D) was determined with one-way analysis of variance (ANOVA) and P-values are from Tukey's honestly significant difference (HSD) post hoc analyses.

purified MAT1A C120S and MAT1A C376S mutants. MAT1A C120S activity displayed no significant reduction in activity in the presence of 100 and 500 μM FA (Fig. 3B), indicating a functional role of Cys120 in FA-sensitivity. By contrast, MAT1A C376S activity displayed comparable activity to that of wild-type (WT) MAT1A with 58% inhibition from 100 μM FA and 68% inhibition from 500 μM FA (Fig. 3B). A MAT2A G120C mutant was created and analyzed with no significant reduction in activity with 100 μM FA and slightly decreased activity with 500 μM FA (fig. S7A), indicating that additional factors around the Gly120 residue (fig. S7B) (*39*) may drive the role of FA modification in MAT2A.

We analyzed MAT1A activity at lower FA concentrations, and even 10 μM FA was sufficient to reduce the apparent catalytic rate constant (k_{cat}) by 22%, with treatments of 25 and 100 μM FA further reducing k_{cat} by 59% and 88%, respectively, showing dose-dependent MAT1A inhibition in a physiologically relevant 10 to 100 μM range (Fig. 3C, fig. S7C, and table S3). FA treatments also reduced substrate binding affinity as reflected by increasing Michaelis constant (K_M) values for FA (fig. S7C and table S3), though to a lesser extent, which shows that MAT1A inhibition by FA is largely k_{cat} driven. We also assessed the selectivity of FA for MAT1A inhibition over hydrogen peroxide (H_2O_2), nitric oxide (NO), and acetaldehyde. There was only modest inhibition of MAT1A upon treatment with 100 μM H_2O_2 (12% reduction) and 100 μM NO (31% reduction) (Fig. 3C, fig. S7D, and table S3) at concentrations that represent a large excess of these species, typically found in the nanomolar range (*18*). We also did not observe significant change in k_{cat} with 100 μM acetaldehyde (fig. S7, E and F, and table S3), showing that MAT1A inhibition is FA selective.

MAT1A-positive, but not MAT2A-positive cells, respond to FA by decreasing cellular SAM biosynthesis

To evaluate the effects of FA-dependent inhibition of MAT1A activity in a cellular context, we used CRISPR-Cas9–mediated knockout of *MAT1A* and *MAT2A* in hepatoblastoma-derived HepG2 cells. HepG2 cells endogenously express both MAT1A and MAT2A, making them ideally suited to create a pair of matched cellular models that express only MAT1A or MAT2A. We successfully created two such cell lines: *MAT2A* knockout (KO) HepG2 cells expressing only MAT1A (MAT1A-positive) and *MAT1A* KO HepG2 cells expressing only MAT2A (MAT2A-positive) (fig. S8A).

Based on the collective data from proteomics and biochemical experiments, we reasoned that FA-dependent inhibition of MAT1A would result in a decrease in cellular SAM levels and methylation of downstream targets, whereas the methylation potential of MAT2A-dependent cells would be insensitive to FA. To test this hypothesis, we measured levels of SAM production for MAT activity in response to varying doses of FA. We applied ethionine supplementation as a MAT substrate to create the more stable cellular product *S*-adenosylethionine (SAE) (*37*), which cannot be used by downstream methyltransferases, resulting in

an intracellular buildup of SAE that could be quantified through mass spectrometry. Cells were also incubated in a one-carbon restriction (1CR) media lacking serine, glycine, and methionine, which resulted in depletion of methyl units (*43*). After validating the use of SAE and 1CR (fig. S8B), results showed a dose-dependent decrease in SAE production in MAT1A-positive cells in response to 0, 200, and 400 μM FA and no change in SAE production in MAT2A-positive cells (Fig. 3D), confirming that MAT2A activity is not sensitive to FA. To determine the biological relevance of MAT1A inhibition on SAM levels, we determined a SAM to *S*-adenosylhomocysteine (SAH) ratio of 23.1 in MAT1A-positive cells compared with a ratio of 20.5 in MAT1A-positive cells with 200 μM FA, 18.4 in HepG2 WT cells, and 7.6 in MAT2A-positive cells (fig. S8B), showing that even when compensatory mechanisms and one-carbon units are available, inhibition by FA and the more extreme full knockout of MAT1A resulted in a decrease in cellular SAM.

Genetic mouse models of chronic FA overload show deficiencies in MAT1A-dependent SAM production and methylation on privileged targets

The collective data identify a reciprocal feedback cycle in which FA exposure and elevation inhibits SAM biosynthesis in a dose-dependent and isoform-specific manner. Given that FA-related diseases arise from chronic rather than acute exposure (*44*–*47*), we sought to test this model in vivo to help decipher the physiological relationships between FA and SAM pools and their effects on downstream methylation pathways (Fig. 4A). We investigated a previously developed *Adh5*$^{-/-}$ mouse that has been used as a model of chronic FA stress (supplementary text) (*10*, *48*, *49*). *Adh5*$^{-/-}$ mice were determined to have elevated FA and N^2-hydroxymethyldeoxyguanosine, a major FA adduct on DNA (*50*), compared with their WT counterparts (*10*, *21*, *22*). FA is detoxified through two clearance mechanisms that rely on ADH5 and ALDH2; the double-KO mice were visibly smaller than individual-KO mice, and the rare survivors did not live longer than 47 weeks (*22*).

We first assessed the ratio of SAM to SAH as a measure of cellular methylation potential (*51*), hypothesizing that FA-dependent inhibition of MAT1A would result in a decreased SAM/SAH ratio in *Adh5*$^{-/-}$ mice liver compared with WT. Metabolomic data quantifying extracted metabolites indeed showed a significant decrease in SAM in *Adh5*$^{-/-}$ liver compared with WT with no change in SAH (Fig. 4B). We next examined if lowered SAM/SAH ratios would be reflected in inhibition of SAM-dependent methylation of DNA and mRNA. However, quantification of 5-methyldeoxycytidine (5mC) as a measure of global DNA methylation (Fig. 4C) and N^6-methyladenosine (m6A) as a measure of global mRNA methylation (Fig. 4D) by mass spectrometry displayed no significant changes between WT and *Adh5*$^{-/-}$ mice liver samples. These results aligned with previous observations that *Adh5*$^{-/-}$ mice are physiologically comparable to their WT counterparts and may be able to sustain methylation through compensatory aldehyde dehydrogenase activity (*22*). In addition to DNA and mRNA, we also examined histone methylation with particular interest in monitoring mono-, di-, and trimethylation of H3K4, K36, and K79 histones that were reported as methyl sinks (*52*) associated with transcriptional activation, elongation (*53*), and chromatin structure (*54*). Indeed, we observed that in histones purified from *Adh5*$^{-/-}$ mice liver, K4 and K79 showed statistically significant decreases in methylation in mono- and dimethylation (Fig. 4E and fig. S9, A and C). By contrast, K9, K27, and K36 methylation and K27 acetylation were not found to be significantly changed in *Adh5*$^{-/-}$ histones (Fig. 4F and fig. S9, B and C). The data are in line with the highly prioritized and compartmentalized nature of one-carbon metabolism pathways (*4*), showing privileged specificity for certain methyl sinks that are affected by chronic elevations in cytosolic FA pools despite a global decrease in SAM.

FA-inhibited SAM production leads to isoform-specific compensatory increased MAT1A expression through genetic and epigenetic mechanisms

During periods of chronic liver stress, such as in hepatocellular carcinoma, an isoform switch from MAT1A to MAT2A leads to an overall reduction in basal levels of SAM production (*36*, *37*). As such, we sought to test how chronic FA elevations in the genetic *Adh5*$^{-/-}$ mouse model may influence overall MAT expression to compensate for the sustained loss of MAT1A activity (Fig. 5A). We first measured MAT1A and MAT2A protein expression levels in *Adh5*$^{-/-}$ mice liver lysate; immunoblotting showed that protein expression for MAT1A was elevated 2.8-fold in *Adh5*$^{-/-}$ liver over WT, but MAT2A expression was not significantly changed (Fig. 5B and fig. S10A). Therefore, in this in vivo context, SAM deficiency caused by chronic FA overload is indeed a result of inhibition of MAT1A activity rather than an isoform switch to MAT2A, with the observed increase in MAT1A expression as a potential compensatory mechanism to combat this deficiency.

To decipher the mechanism by which MAT1A expression was altered by chronic FA elevation, we quantified the normalized expression change of mRNA transcripts for *Adh5*$^{-/-}$ compared with that of WT. We found that the trends for both *Mat1a* and *Mat2a* mRNA levels paralleled the trends in protein expression (Fig. 5C), indicating that compensation is mediated at the transcriptional level. In this context, the MAT1A-to-MAT2A expression switch has been reported to be regulated by DNA methylation, where hypo- or unmethylated MAT1A and MAT2A promoters are known to induce expression of the respective isoform (*40*, *55*, *56*). To investigate methylation of *Mat1a* and *Mat2a* sites, targeted next-generation bisulfite sequencing was analyzed in the mouse *Mat1a* and *Mat2a* promoters and gene bodies to identify changes in methylation that potentially regulate enzyme expression (*57*–*59*). We observed select CpG sites in the *Mat1a* promoter that showed reduced methylation in *Adh5*$^{-/-}$ compared with that of WT liver (Fig. 5D). The observed decreases in methylation of the *Mat1a* promoter with concomitant increases in mRNA transcripts and protein expression are in line with reports that show that hypomethylation of the *Mat1a* promoter results in an increase in expression (*55*, *56*). By contrast, we did not observe FA-dependent methylation changes of CpG sites in *Mat2a* (Fig. 5E). These results support the model that SAM deficiency upon FA overload is regulated by MAT1A activity in an isoform-specific manner, with DNA hypomethylation as one of the potential compensatory mechanisms that elevate MAT1A expression and thus help maintain SAM levels in response to this deficiency.

To further support whether the observed changes in *Mat1a*-promoter CpG methylation and MAT1A expression were direct consequences of elevated FA levels, rather than disruptions from formate and other one-carbon metabolite sources upon *Adh5*$^{-/-}$, we used cell models with acute FA treatment to monitor CpG methylation and MAT1A and MAT2A expression. We exposed HepG2 cells to 200 μM FA and analyzed CpG methylation and protein expression at 2- to 10-hour timepoints relative to a 10-hour vehicle control. Similar to what was observed in the *Adh5*$^{-/-}$ mouse model, we observed an increase in MAT1A expression (0.047-fold change per hour) and mRNA transcripts (0.026-fold change per hour) over time, but MAT2A expression and transcripts remained largely unchanged (fig. S10B and fig. S11, A and B). A similar experiment with 5-mM SAM supplementation displayed no significant change in MAT1A expression (fig. S11, C and D), suggesting that the increase in MAT1A expression from FA treatment is SAM-dependent. Analysis of targeted bisulfite sequencing of CpG sites with higher than 2% methylation revealed that, aside from CpG number −15, the assayed *MAT1A* CpG sites were either unchanged or had a negative rate of methylation percent change per hour over the 10-hour FA treatment; that of *MAT2A* sites instead increased over time (fig. S11, E and F). Together, the cell and mouse model data are consistent with FA overload triggering a compensatory, isoform-specific

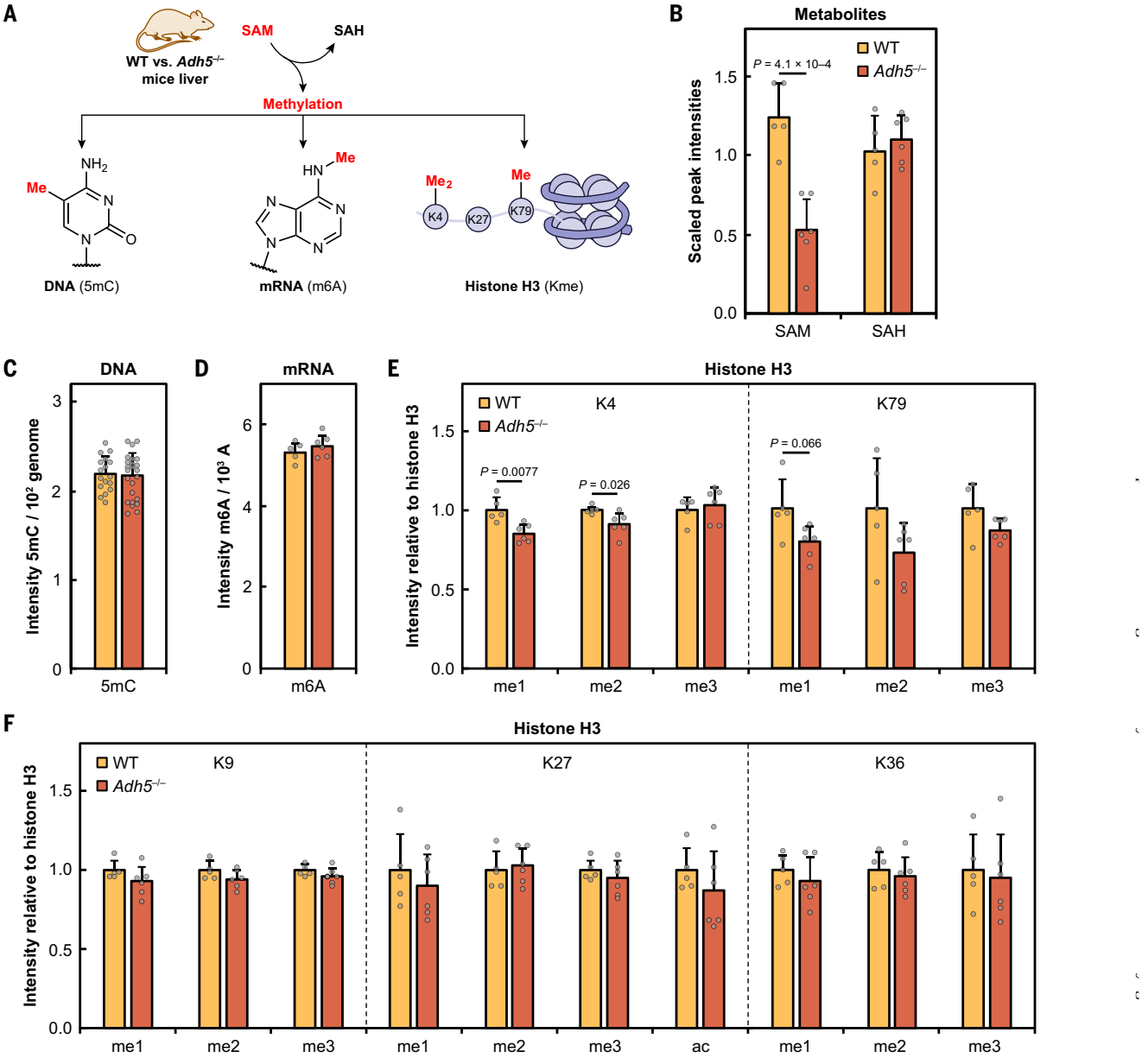


Fig. 4. Genetic mouse model of chronic FA overload has reduced methylation potential that specifically targets histone methyl sinks. (**A**) Schematic of SAM and its downstream methylation that were measured in WT and *Adh5*$^{-/-}$ mice liver. (**B**) SAM and SAH measured by mass spectrometry in WT and *Adh5*$^{-/-}$ liver. (**C**) Global DNA methylation measured by 5mC and normalized by deoxycytidine (dC). (**D**) Global RNA methylation measured by m6A and normalized by adenosine (A). (**E**) Histone H3 methylation measured by immunoblotting for K4 and K79 mono- (me1), di- (me2), and trimethylation (me3). (**F**) Histone H3 methylation measured by immunoblotting for K9, K27, and K36 mono-, di-, and trimethylation, and K27 acetylation (ac). All histone blots were normalized by total histone H3. WT (n = 5; yellow) and *Adh5*$^{-/-}$ (n = 6; orange) of biological replicates. Error bars represent SD for all graphs. Statistical significance was determined with two-tailed t test.

increase in MAT1A expression through genetic and epigenetic regulation mechanisms.

Chronic FA elevation alters genome-wide methylation and transcriptional regulation

We next explored effects of chronic FA elevation on genome-wide gene regulation. Indeed, in this context it has been shown that SAM treatment in HepG2 cells leads to differential hypo- and hypermethylation of select DNA CpG sites (*60*). As such, we performed genome-wide CpG methylation analysis in liver samples from *Adh5*$^{-/-}$ mice compared with those of WT controls. *Mat1a* and *Mat2a* CpG DNA methylation were analyzed in the genome-wide dataset to support the results from the targeted bisulfite sequencing. We observed a trend of hypomethylation in *Adh5*$^{-/-}$ samples in four of the six analyzed *Mat1a* CpG sites (Fig. 5F). No change was observed for the five CpG sites for *Mat2a* (fig. S12A). Similar to a previous report on SAM-dependent DNA methylation changes (*60*), we observed both hypo- and hypermethylation of CpG sites in *Adh5*$^{-/-}$ compared with those of WT (Fig. 5G, fig. S12B, and table S4). Pathway analysis showed that there was selective hypomethylation of CpGs present in metabolic genes involved in alcoholic liver disease and fatty acid metabolism and affecting genes containing the binding-site motifs of c-Myc, ZF5, and E2F1 in *Adh5*$^{-/-}$ mice (Fig. 5H and fig. S12C). In addition, a gain of methylation was observed in genes involved in propanoate metabolism and in oxidoreductase activity on alcohol group of donors (Fig. 5I and fig. S12D). The genes containing binding-site motifs for FOXO3a

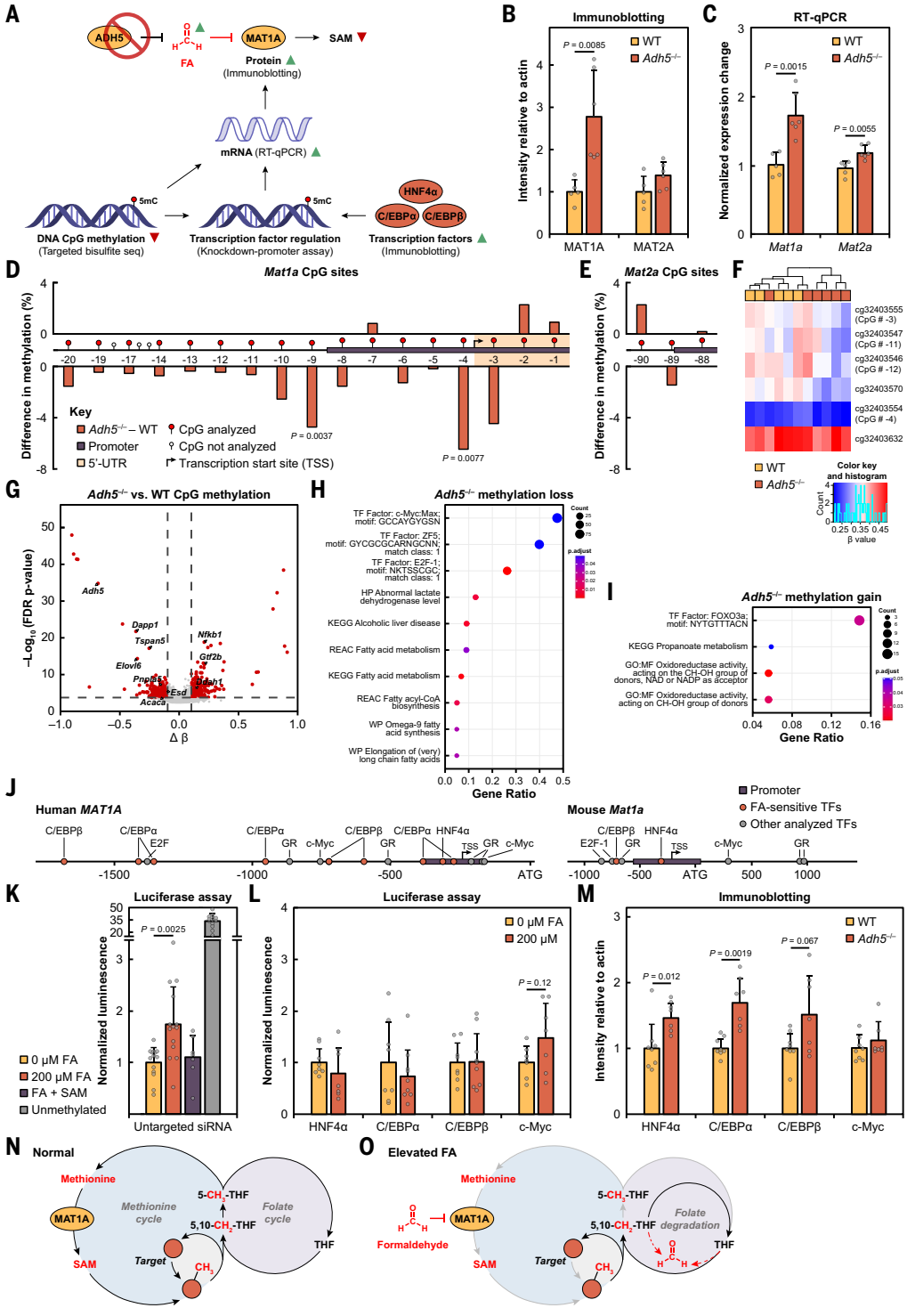


Fig. 5. Chronic elevations in FA induce compensatory increases in MAT1A expression by decreasing promoter methylation. (**A**) Schematic and summary of results of MAT1A regulation in response to elevated FA in *Adh5*$^{-/-}$ mice liver. (**B**) MAT1A and MAT2A protein expression measured by immunoblotting for WT (n = 5; yellow) and *Adh5*$^{-/-}$ liver (n = 6; orange) of biological replicates. (**C**) *Mat1a* and *Mat2a* mRNA transcript levels measured by RT-qPCR. (**D** to **E**) Difference in percent methylation of (D) *Mat1a* and (E) *Mat2a* DNA CpG sites between *Adh5*$^{-/-}$ and WT liver. The x axis shows schematic of *Mat1a* and *Mat2a* promoters for CpG sites with >2% methylation. Sites are numbered sequentially relative to the ATG site. P values were determined with two-way t test. (**F**) Heatmap of CpG β values for MAT1A sites analyzed in the genome-wide CpG analysis. β values represent percent methylation from zero (fully unmethylated) to one (fully methylated). (**G**) Volcano plot of the delta between *Adh5*$^{-/-}$ compared to WT for genome-wide CpG β values. One-carbon and fatty acid metabolism gene hits of interest are labeled. (**H**) Pathway enrichment analysis of methylation loss in *Adh5*$^{-/-}$. (**I**) Pathway enrichment analysis of methylation gain in *Adh5*$^{-/-}$. (**J**) Schematic of human *MAT1A* and mouse *Mat1a* promoter with TFs studied in this work labeled. TFs that were deemed as important for *MAT1A* promoter sensitivity for FA are in orange. (**K**) Luciferase assay of HepG2 human *MAT1A* methylated and unmethylated promoters with

0 or 200 μM FA and 5 mM SAM treatments. Data are normalized to untreated control. 0 and 200 μM FA treatment (n = 14) technical cell replicates from four experiments, FA with SAM treatment (n = 7) technical cell replicates from two experiments, and unmethylated promoter (n = 13) technical cell replicates from four experiments. (**L**) Luciferase assay of HepG2 siRNA-mediated TF knockdown with human *MAT1A* methylated promoter with 0 or 200 μM FA. Data are normalized to each untreated control. HepG2 (n = 7 to 9) of technical cell replicates from two experiments. (**M**) Expression of several MAT1A-associated TFs in WT (n = 8) versus *Adh5*$^{-/-}$ (n = 7) of biological replicates. Error bars represent SD for all graphs. Statistical significance was determined with a two-tailed t test. (**N**) Proposed model of normal one-carbon metabolism in cells through MAT1A-catalyzed production of SAM and demethylation reactions, preserving the carbon unit through folate-mediated synthesis of methionine. (**O**) Proposed model of FA-dependent regulation of one-carbon metabolism in cells under situations of FA overload, showing decrease of SAM biosynthesis from isoform-specific FA-inhibition of MAT1A (gray arrows), disruption of the folate cycle due to hypermethioninemia, and further FA elevation from spontaneous folate degradation (dashed red arrows). THF, tetrahydrofolate.

presented a significant gain of methylation in *Adh5*$^{-/-}$ compared with WT. Remarkably, most of the genes showing loss or gain of methylation are involved in metabolic pathways, which suggests that FA might drive a metabolic reprogramming through selectively altering CpG methylation. The expression of genes controlled by c-Myc or FOXO3a might be significantly altered by endogenous FA. Subsequently, analysis of transcription factor (TF) binding profiles with the JASPAR database (*61*) revealed 10 TF binding profiles with significant enrichment ($P < 0.01$) among the hypomethylated CpG sites and 37 TF binding profiles among the hypermethylated CpG sites (fig. S13, A and B, and table S5). Focusing on CpG sites located within promoter regions identified two TF binding profiles (CEBPB and NFIX) for CpG hypomethylation and three TF binding profiles (ZNF707, PAX3, and FOSB:JUNB) for CpG hypermethylation (fig. S13, C and D, and table S5), exhibiting TFs with potential key roles in FA-dependent gene regulation.

To provide functional support for the effect of FA on MAT1A expression through regulation of the *MAT1A* promoter (Fig. 5J), we performed a luciferase assay with a *MAT1A* promoter–luciferase plasmid (*55*). In vitro methylation was performed to obtain a pair of fully unmethylated active and fully methylated inactive *MAT1A* plasmids. HepG2 cells were transfected with methylated *MAT1A* promoter–luciferase plasmid and luminescence was measured, with the unmethylated *MAT1A* promoter plasmid as a positive control for promoter activity. Cells treated with 200 μM FA for 10 hours showed a 1.7-fold-change increase in luciferase activity compared with untreated samples (Fig. 5K). The observed increase in FA-triggered promoter activity was negated by 5 mM SAM, supporting the idea that the effect of FA is SAM-dependent.

Because the changes in *Mat1a* promoter hypomethylation are modest and alone likely do not fully explain the large compensatory increase in MAT1A expression upon FA exposure, we investigated other potential regulatory mechanisms. To this end, MAT1A is known to be regulated by several TFs, including but not limited to hepatocyte nuclear factor 4-alpha (HNF4α), CCAAT/enhancer-binding protein alpha (C/EBPα), CCAAT/enhancer-binding protein beta (C/EBPβ), glucocorticoid receptor (GR), Myc proto-oncogene protein (c-Myc), and transcription factor E2F1 (E2F1) (Fig. 5J) (*40*). To determine the potential contributions of these TFs on FA-dependent regulation of MAT1A expression, we applied the luciferase *MAT1A* promoter assay after individual knockdown of the six TFs or an untargeted control small interfering RNA (siRNA) (fig. S14A). Upon knockdown of HNF4α, C/EBPα, and C/EBPβ, we observed that promoter activity was insensitive to FA treatment (Fig. 5L), showing that these TFs are involved in FA-dependent MAT1A regulation. C/EBPβ was of particular interest as its binding profile was enriched in promoters in our TFs analysis (fig. S13C). By contrast, upon knockdown of c-Myc, increases in FA-dependent promoter activity were comparable to that of untargeted siRNA control (Fig. 5, K and L).

The functional cellular assays are supported by immunoblotting data, showing that levels of HNF4α, C/EBPα, and C/EBPβ proteins, but not c-Myc, are higher in *Adh5*$^{-/-}$ mice relative to WT (Fig. 5M and fig. S14B). These results are in line with reports showing that these TFs can positively regulate MAT1A expression (*57*, *62*). Although its putative binding motif was found to be enriched in differentially methylated positions in our genome-wide CpG analysis, c-Myc expression was unchanged in *Adh5*$^{-/-}$ mice. The TFs GR and E2F1 were also analyzed (fig. S14, B to D), but their expression increases were not correlated with significance in FA-sensitivity, suggesting that more complex factors, including localization, binding, and activity, are in play. There may also be a difference in mechanism between human hepatoblastoma HepG2 cells and normal mice liver, as the human and murine *MAT1A* promoters differ (Fig. 5J). Together with our data that shows *MAT1A* promoter sensitivity to FA with these TFs, the observed increases of HNF4α, C/EBPα, and C/EBPβ expression in the *Adh5*$^{-/-}$ mouse model of chronic FA elevation establish their participation in FA-SAM–dependent regulation of MAT1A expression.

Conclusions

Activity-based protein profiling of FA-reactive cysteine residues revealed that FA is not an indiscriminate electrophile, but instead serves as a regulatory posttranslational modifier that can target a privileged class of hyperreactive cysteine residues in the proteome. By focusing on its contributions in one-carbon metabolism, we characterized an isoform-specific inhibition of MAT1A activity by FA at a key, conserved cysteine site. We identified a biochemical feedback cycle between two ubiquitous one-carbon units, in which FA can regulate the biosynthetic production of SAM, the major methyl donor of the cell. Using an *Adh5*$^{-/-}$ mouse model of chronic FA overload, we established the physiological significance of this one-carbon crosstalk by observing that FA-dependent inhibition of SAM biosynthesis decreases the methylation potential of the cell in a highly selective manner, despite a global decrease in SAM levels. Specifically, we observed methylation changes in privileged K4 and K79 histone methylation sites as methyl sinks, rather than ubiquitous, global changes to DNA, RNA, and histone methylation targets. Additionally, we uncovered a compensatory feedback pathway for FA-dependent SAM deficiency, in which chronic and acute FA overload result in selective increase in MAT1A expression. A FA-MAT1A-SAM feedback cycle is mediated through FA-dependent MAT1A inhibition and a combination of epigenetic regulation by means of altered methylation at select transcription-factor binding sites and increased expression of select TFs to trigger with compensatory up-regulation of MAT1A expression.

The collective data support a model where in healthy cells, when FA homeostasis is tightly maintained, the methyl unit on methionine is converted to SAM for use in downstream methyltransferase writers that install one-carbon posttranslational modifications on proteins, nucleic acids, and other metabolic substrates (Fig. 5N) (*63*). In turn, a number of FA-generating demethylase enzyme erasers for these posttranslational modifications, such as lysine-specific demethylase, possess folate binding pockets that sequester the released FA into 5,10-methylenetetrahydrofolate, which is then converted into 5-methyltetrahydrofolate to methylate homocysteine and produce methionine (*64*). As such, cells achieve a biochemical conservation of one-carbon units within the one-carbon cycle by balancing SAM biosynthesis and methylation or demethylation reactions. In cells with elevated levels of FA owing to toxic exposure, metabolic imbalances, and/or disease progression, dysregulation of carbon units stems from the decreased biosynthesis of SAM through MAT1A regulation and results in hypomethylation of

downstream disease targets (Fig. 5O). This dysregulation can lead to epigenetic dysregulation and genome instability. Indeed, hypermethioninemia occurs from decreased MAT1A activity (*65*), leading to a buildup of tetrahydrofolate intermediates that we have previously shown to spontaneously degrade to release FA (*10*). In turn, this elevated FA production further inhibits the biosynthesis of SAM, leading to a global depletion of one-carbon units in the cell. Owing to the universal nature of one-carbon metabolism in all organisms and the central roles of SAM-dependent biochemical processes in the cell, the identification of a molecular interplay between one-carbon FA and SAM biosynthesis sets the stage for further investigations of how these reactive carbon species contribute to biological function.

Materials and methods summary

Full details of this study's methods are described in the supplementary materials. Methods are briefly listed here. isoTOP-ABPP and proteomics data analyses were performed as previously described (*66*) with mouse liver tissue lysate and 500-μM FA treatment. Site-directed mutagenesis was performed on MAT1A and MAT2A WT plasmids to obtain MAT1A C120S, MAT1A C376S, and MAT2A G120C mutants. His_6-tagged MAT proteins were purified for biochemical analyses and protein crystallography; the structure was processed and refined using HKL2000 (*67*), Phenix (*68*), and COOT (*69*). FA modifications on MAT1A peptides were analyzed by Q Exactive mass spectrometry. MAT activity assays were performed with FA treatment, addition of methionine and ATP substrates, quenching over multiple timepoints, and analysis with Enzkin (*70*) or Prism. CRISPR-mediated KO of *MAT1A* and *MAT2A* was performed in HepG2 human hepatoblastoma cells to demonstrate isoform-specificity of FA treatment through the measurement of SAE as a stable proxy of SAM. To quantify methylation potential in HepG2 WT and KO cell lines upon FA treatment, SAM and SAH were extracted by adapting a previously reported protocol (*71*). Metabolites were measured by triple quadrupole LC-MS. An endogenous elevated FA model, $Adh5^{-/-}$ mice, was then studied in comparison to WT. Untargeted metabolomics was performed by Metabolon in mouse liver tissue as previously described (*72*). 5mC DNA methylation in the same cohort of mice as Mulderrig *et al.* (*49*) and m6A mRNA methylation were measured by mass spectrometry. Histone methylation was analyzed in purified histones by immunoblotting with modification-specific antibodies, with specificity insight from the histone antibodies database (*73*). MAT protein expression was analyzed by immunoblotting, and mRNA transcripts were analyzed by reverse transcription quantitative polymerase chain reaction (RT-qPCR) and the $2^{-\Delta\Delta CT}$ quantification method (*74*). CpG DNA methylation was analyzed by targeted bisulfite sequencing, and global DNA methylation was analyzed with Infinium Mouse Methylation BeadChip Array. Enrichment of transcription-factor binding profiles was analyzed with JASPAR database (*61*, *75*). TF siRNA knockdown in HepG2 was performed and followed by transfection of luciferase MAT1A promoter plasmid, provided by the Lu lab (*55*). Luciferase assay was performed to demonstrate the effects of FA on MAT1A promoter activity.

REFERENCES AND NOTES

1. J. W. Locasale, Serine, glycine and one-carbon units: Cancer metabolism in full circle. *Nat. Rev. Cancer* **13**, 572–583 (2013). doi: 10.1038/nrc3557; pmid: 23822983
2. G. S. Ducker, J. D. Rabinowitz, One-carbon metabolism in health and disease. *Cell Metab.* **25**, 27–42 (2017). doi: 10.1016/j.cmet.2016.08.009; pmid: 27641100
3. S. C. Lu, J. M. Mato, S-adenosylmethionine in liver health, injury, and cancer. *Physiol. Rev.* **92**, 1515–1542 (2012). doi: 10.1152/physrev.00047.2011; pmid: 23073625
4. A. S. Tibbetts, D. R. Appling, Compartmentalization of Mammalian folate-mediated one-carbon metabolism. *Annu. Rev. Nutr.* **30**, 57–81 (2010). doi: 10.1146/annurev.nutr.012809.104810; pmid: 20645850
5. X. Gu *et al.*, SAMTOR is an *S*-adenosylmethionine sensor for the mTORC1 pathway. *Science* **358**, 813–818 (2017). doi: 10.1126/science.aao3265; pmid: 29123071
6. M. Noureddin, J. M. Mato, S. C. Lu, Nonalcoholic fatty liver disease: Update on pathogenesis, diagnosis, treatment and the role of S-adenosylmethionine. *Exp. Biol. Med.* **240**, 809–820 (2015). doi: 10.1177/1535370215579161; pmid: 25873078
7. M. Yang, K. H. Vousden, Serine and one-carbon metabolism in cancer. *Nat. Rev. Cancer* **16**, 650–662 (2016). doi: 10.1038/nrc.2016.81; pmid: 27634448
8. T. Gerken *et al.*, The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469–1472 (2007). doi: 10.1126/science.1151710; pmid: 17991826
9. R. J. Hopkinson, C. J. Schofield, Deciphering functions of intracellular formaldehyde: Linking cancer and aldehyde metabolism. *Biochemistry* **57**, 904–906 (2018). doi: 10.1021/acs.biochem.7b01304; pmid: 29368521
10. G. Burgos-Barragan *et al.*, Mammals divert endogenous genotoxic formaldehyde into one-carbon metabolism. *Nature* **548**, 549–554 (2017). doi: 10.1038/nature23481; pmid: 28813411
11. C. Yi, C.-G. Yang, C. He, A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc. Chem. Res.* **42**, 519–529 (2009). doi: 10.1021/ar800178j; pmid: 19852088
12. L. J. Walport, R. J. Hopkinson, C. J. Schofield, Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **16**, 525–534 (2012). doi: 10.1016/j.cbpa.2012.09.015; pmid: 23063108
13. J. Ohata, K. J. Bruemmer, C. J. Chang, Activity-based sensing methods for monitoring the reactive carbon species carbon monoxide and formaldehyde in living systems. *Acc. Chem. Res.* **52**, 2841–2848 (2019). doi: 10.1021/acs.accounts.9b00386; pmid: 31487154
14. T. F. Brewer, C. J. Chang, An Aza-Cope Reactivity-Based Fluorescent Probe for Imaging Formaldehyde in Living Cells. *J. Am. Chem. Soc.* **137**, 10886–10889 (2015). doi: 10.1021/jacs.5b05340; pmid: 26306005
15. K. J. Bruemmer *et al.*, Development of a General Aza-Cope Reaction Trigger Applied to Fluorescence Imaging of Formaldehyde in Living Cells. *J. Am. Chem. Soc.* **139**, 5338–5350 (2017). doi: 10.1021/jacs.6b12460; pmid: 28375637
16. T. F. Brewer, G. Burgos-Barragan, N. Wit, K. J. Patel, C. J. Chang, A 2-aza-Cope reactivity-based platform for ratiometric fluorescence imaging of formaldehyde in living cells. *Chem. Sci.* **8**, 4073–4081 (2017). doi: 10.1039/C7SC00748E; pmid: 28580121
17. K. J. Bruemmer, O. Green, T. A. Su, D. Shabat, C. J. Chang, Chemiluminescent Probes for Activity-Based Sensing of Formaldehyde Released from Folate Degradation in Living Mice. *Angew. Chem. Int. Ed.* **130**, 7630–7634 (2018). doi: 10.1002/ange.201802143; pmid: 29635731
18. S. Parvez, M. J. C. Long, J. R. Poganik, Y. Aye, Redox signaling by reactive electrophiles and oxidants. *Chem. Rev.* **118**, 8798–8888 (2018). doi: 10.1021/acs.chemrev.7b00698; pmid: 30148624
19. Y. Wei *et al.*, Simultaneous determination of seven endogenous aldehydes in human blood by headspace gas chromatography-mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **1118–1119**, 85–92 (2019). doi: 10.1016/j.jchromb.2019.04.027; pmid: 31030105
20. Z. Tong *et al.*, Tumor tissue-derived formaldehyde and acidic microenvironment synergistically induce bone cancer pain. *PLOS ONE* **5**, e10234 (2010). doi: 10.1371/journal.pone.0010234; pmid: 20422007
21. L. B. Pontel *et al.*, Endogenous formaldehyde is a hematopoietic stem cell genotoxin and metabolic carcinogen. *Mol. Cell* **60**, 177–188 (2015). doi: 10.1016/j.molcel.2015.08.020; pmid: 26412304
22. F. A. Dingler *et al.*, Two Aldehyde Clearance Systems Are Essential to Prevent Lethal Formaldehyde Accumulation in Mice and Humans. *Mol. Cell* **80**, 996–1012.e9 (2020). doi: 10.1016/j.molcel.2020.10.012; pmid: 33147438
23. B. Metz *et al.*, Identification of formaldehyde-induced modifications in proteins: Reactions with model peptides. *J. Biol. Chem.* **279**, 6235–6243 (2004). doi: 10.1074/jbc.M310752200; pmid: 14638685
24. J. J. A. G. Kamps, R. J. Hopkinson, C. J. Schofield, T. D. W. Claridge, How formaldehyde reacts with amino acids. *Commun. Chem.* **2**, 126 (2019). doi: 10.1038/s42004-019-0224-2
25. E. Weerapana *et al.*, Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790–795 (2010). doi: 10.1038/nature09472; pmid: 21085121
26. E. Weerapana, A. E. Speers, B. F. Cravatt, Tandem orthogonal proteolysis-activity-based protein profiling (TOP-ABPP)—A general method for mapping sites of probe modification in proteomes. *Nat. Protoc.* **2**, 1414–1425 (2007). doi: 10.1038/nprot.2007.194; pmid: 17545978
27. S. M. Couvertier, Y. Zhou, E. Weerapana, Chemical-proteomic strategies to investigate cysteine posttranslational modifications. *Biochim. Biophys. Acta* **1844**, 2315–2330 (2014). doi: 10.1016/j.bbapap.2014.09.024; pmid: 25291386
28. J. S. Coukos, C. W. Lee, K. S. Pillai, K. J. Liu, R. E. Moellering, Widespread, Reversible Cysteine Modification by Methylglyoxal Regulates Metabolic Enzyme Function. *ACS Chem. Biol.* **18**, 91–101 (2023). doi: 10.1021/acschembio.2c00727; pmid: 36562291
29. Y.-H. Huang, E.-P. I. Chiang, Dietary Glycine Alters One-Carbon Metabolic kinetics in vivo. *FASEB J.* **30**, 272–274 (2016).
30. S. Waris, A. Patel, A. Ali, R. Mahmood, Acetaldehyde-induced oxidative modifications and morphological changes in isolated human erythrocytes: An in vitro study. *Environ. Sci. Pollut. Res. Int.* **27**, 16268–16281 (2020). doi: 10.1007/s11356-020-08044-4; pmid: 32124282
31. J. Komoto, T. Yamada, Y. Takata, G. D. Markham, F. Takusagawa, Crystal structure of the S-adenosylmethionine synthetase ternary complex: A novel catalytic mechanism of S-adenosylmethionine synthesis from ATP and Met. *Biochemistry* **43**, 1821–1831 (2004). doi: 10.1021/bi035611t; pmid: 14967023
32. E. Sánchez-Góngora *et al.*, Interaction of liver methionine adenosyltransferase with hydroxyl radical. *FASEB J.* **11**, 1013–1019 (1997). doi: 10.1096/fasebj.11.12.9337154; pmid: 9337154
33. M. A. Avila *et al.*, Regulation of rat liver S-adenosylmethionine synthetase during septic shock: Role of nitric oxide. *Hepatology* **25**, 391–396 (1997). doi: 10.1002/hep.510250222; pmid: 9021952
34. I. Pérez-Mato, C. Castro, F. A. Ruiz, F. J. Corrales, J. M. Mato, Methionine adenosyltransferase S-nitrosylation is regulated by the basic and acidic amino acids surrounding the target thiol. *J. Biol. Chem.* **274**, 17075–17079 (1999). doi: 10.1074/jbc.274.24.17075; pmid: 10358060
35. S. C. Lu, Methionine adenosyltransferase and liver disease: It's all about SAM. *Gastroenterology* **114**, 403–407 (1998). doi: 10.1016/S0016-5085(98)70494-9; pmid: 9453503
36. M. Frau, F. Feo, R. M. Pascale, Pleiotropic effects of methionine adenosyltransferases deregulation as determinants of liver cancer progression and prognosis. *J. Hepatol.* **59**, 830–841 (2013). doi: 10.1016/j.jhep.2013.04.031; pmid: 23665184
37. C. L. Quinlan *et al.*, Targeting S-adenosylmethionine biosynthesis with a novel allosteric inhibitor of Mat2A.

Nat. Chem. Biol. **13**, 785–792 (2017). doi: 10.1038/nchembio.2384; pmid: 28553945
38. L. Y. Maldonado, D. Arsene, J. M. Mato, S. C. Lu, Methionine adenosyltransferases in cancers: Mechanisms of dysregulation and implications for therapy. *Exp. Biol. Med.* **243**, 107–117 (2018). doi: 10.1177/1535370217740860; pmid: 29141455
39. N. Shafqat *et al*., Insight into S-adenosylmethionine biosynthesis from the crystal structures of the human methionine adenosyltransferase catalytic and regulatory subunits. *Biochem. J.* **452**, 27–36 (2013). doi: 10.1042/BJ20121580; pmid: 23425511
40. J. M. Mato, F. J. Corrales, S. C. Lu, M. A. Avila, S-Adenosylmethionine: A control switch that regulates liver function. *FASEB J.* **16**, 15–26 (2002). doi: 10.1096/fj.01-0401rev; pmid: 11772932
41. K. Ramani, J. M. Mato, S. C. Lu, Role of methionine adenosyltransferase genes in hepatocarcinogenesis. *Cancers* **3**, 1480–1497 (2011). doi: 10.3390/cancers3021480; pmid: 24212770
42. H. D. Heck, E. L. White, M. Casanova-Schmitz, Determination of formaldehyde in biological tissues by gas chromatography/mass spectrometry. *Biomed. Mass Spectrom.* **9**, 347–353 (1982). doi: 10.1002/bms.1200090808; pmid: 7126766
43. O. D. K. Maddocks, C. F. Labuschagne, P. D. Adams, K. H. Vousden, Serine metabolism supports the methionine cycle and DNA/RNA methylation through de novo ATP synthesis in cancer cells. *Mol. Cell* **61**, 210–221 (2016). doi: 10.1016/j.molcel.2015.12.014; pmid: 26774282
44. L. Zhang *et al*., Occupational exposure to formaldehyde, hematotoxicity, and leukemia-specific chromosome changes in cultured myeloid progenitor cells. *Cancer Epidemiol. Biomarkers Prev.* **19**, 80–88 (2010). doi: 10.1158/1055-9965.EPI-09-0762; pmid: 20056626
45. R. V. Phillips *et al*., Epigenome-wide association studies of occupational exposure to benzene and formaldehyde. *Epigenetics* **17**, 2259–2277 (2022). doi: 10.1080/15592294.2022.2115604; pmid: 36017556
46. M. E. Andersen *et al*., Formaldehyde: Integrating dosimetry, cytotoxicity, and genomics to understand dose-dependent transitions for an endogenous compound. *Toxicol. Sci.* **118**, 716–731 (2010). doi: 10.1093/toxsci/kfq303; pmid: 20884683
47. Q. Liu *et al*., Effects of long-term low-dose formaldehyde exposure on global genomic hypomethylation in 16HBE cells. *Toxicol. Lett.* **205**, 235–240 (2011). doi: 10.1016/j.toxlet.2011.05.1039; pmid: 21745553
48. L. Liu *et al*., Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**, 617–628 (2004). doi: 10.1016/S0092-8674(04)00131-X; pmid: 14980227
49. L. Mulderrig *et al*., Aldehyde-driven transcriptional stress triggers an anorexic DNA damage response. *Nature* **600**, 158–163 (2021). doi: 10.1038/s41586-021-04133-7; pmid: 34819667
50. B. C. Moeller *et al*., Determination of N2-hydroxymethyl-dG adducts in the nasal epithelium and bone marrow of nonhuman primates following 13CD2-formaldehyde inhalation exposure. *Chem. Res. Toxicol.* **24**, 162–164 (2011). doi: 10.1021/tx1004166; pmid: 21222454
51. D. R. Hoffman, D. W. Marion, W. E. Cornatzer, J. A. Duerre, S-Adenosylmethionine and S-adenosylhomocystein metabolism in isolated rat liver. Effects of L-methionine, L-homocystein, and adenosine. *J. Biol. Chem.* **255**, 10822–10827 (1980). doi: 10.1016/S0021-9258(19)70381-0; pmid: 7430157
52. C. Ye, B. M. Sutter, Y. Wang, Z. Kuang, B. P. Tu, A Metabolic Function for Phospholipid and Histone Methylation. *Mol. Cell* **66**, 180–193.e8 (2017). doi: 10.1016/j.molcel.2017.02.026; pmid: 28366644
53. T. Kouzarides, Chromatin modifications and their function. *Cell* **128**, 693–705 (2007). doi: 10.1016/j.cell.2007.02.005; pmid: 17320507
54. J. Choi, Z. Y. Ryoo, D.-H. Cho, H.-S. Lee, H.-Y. Ryu, Trans-tail regulation-mediated suppression of cryptic transcription. *Exp. Mol. Med.* **53**, 1683–1688 (2021). doi: 10.1038/s12276-021-00711-x; pmid: 34845331
55. M. L. Tomasi, T. W. H. Li, M. Li, J. M. Mato, S. C. Lu, Inhibition of human methionine adenosyltransferase 1A transcription by coding region methylation. *J. Cell. Physiol.* **227**, 1583–1591 (2012). doi: 10.1002/jcp.22875; pmid: 21678410
56. H. Yang *et al*., Role of promoter methylation in increased methionine adenosyltransferase 2A expression in human liver cancer. *Am. J. Physiol. Gastrointest. Liver Physiol.* **280**, G184–G190 (2001). doi: 10.1152/ajpgi.2001.280.2.G184 pmid: 11208539
57. R. Ikeda *et al*., Involvement of CCAAT/enhancer binding protein-β (C/EBPβ) in epigenetic regulation of mouse methionine adenosyltransferase 1A gene expression. *Int. J. Biochem. Cell Biol.* **40**, 1956–1969 (2008). doi: 10.1016/j.biocel.2008.02.004; pmid: 18346930
58. W. J. Kent *et al*., The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002). doi: 10.1101/gr.229102; pmid: 12045153
59. K. L. Howe *et al*., Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021). doi: 10.1093/nar/gkaa942; pmid: 33137190
60. Y. Wang, Z. Sun, M. Szyf, S-adenosyl-methionine (SAM) alters the transcriptome and methylome and specifically blocks growth and invasiveness of liver cancer cells. *Oncotarget* **8**, 111866–111881 (2017). doi: 10.18632/oncotarget.22942; pmid: 29340097
61. J. A. Castro-Mondragon *et al*., JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50** (D1), D165–D173 (2022). doi: 10.1093/nar/gkab1113; pmid: 34850907
62. Q. Xu *et al*., HNF4α regulates sulfur amino acid metabolism and confers sensitivity to methionine restriction in liver cancer. *Nat. Commun.* **11**, 3978 (2020). doi: 10.1038/s41467-020-17818-w; pmid: 32770044
63. A.-W. Struck, M. L. Thompson, L. S. Wong, J. Micklefield, S-adenosyl-methionine-dependent methyltransferases: Highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *ChemBioChem* **13**, 2642–2655 (2012). doi: 10.1002/cbic.201200556; pmid: 23180741
64. Z. Luka, F. Moss, L. V. Loukachevitch, D. J. Bornhop, C. Wagner, Histone demethylase LSD1 is a folate-binding protein. *Biochemistry* **50**, 4750–4756 (2011). doi: 10.1021/bi200247b; pmid: 21510664
65. S. H. Mudd *et al*., Isolated hypermethioninemia: Measurements of S-adenosylmethionine and choline. *Metabolism* **49**, 1542–1547 (2000). doi: 10.1053/meta.2000.18521; pmid: 11145114
66. C. Y.-S. Chung *et al*., Covalent targeting of the vacuolar H^+-ATPase activates autophagy via mTORC1 inhibition. *Nat. Chem. Biol.* **15**, 776–785 (2019). doi: 10.1038/s41589-019-0308-4; pmid: 31285595
67. Z. Otwinowski, W. Minor, "Processing of X-ray diffraction data collected in oscillation mode" in *Methods in enzymology* (Elsevier, 1997), vol. 276, pp. 307–326.
68. P. V. Afonine *et al*., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012). doi: 10.1107/S0907444912001308; pmid: 22505256
69. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004). doi: 10.1107/S0907444904019158; pmid: 15572765
70. G. Cardillo, Enzkin: a tool to estimate Michaelis-Menten kinetic parameters (2010); https://www.mathworks.com/matlabcentral/fileexchange/26653-enzkin.
71. M. Kořínek *et al*., Quantification of homocysteine-related metabolites and the role of betaine-homocysteine S-methyltransferase in HepG2 cells. *Biomed. Chromatogr.* **27**, 111–121 (2013). doi: 10.1002/bmc.2755; pmid: 22653757
72. A. M. Evans, C. D. DeHaven, T. Barrett, M. Mitchell, E. Milgram, Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81**, 6656–6667 (2009). doi: 10.1021/ac901536h; pmid: 19624122
73. S. B. Rothbart *et al*., An interactive database for the assessment of histone antibody specificity. *Mol. Cell* **59**, 502–511 (2015). doi: 10.1016/j.molcel.2015.06.022; pmid: 26212453
74. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C(T))$ Method. *Methods* **25**, 402–408 (2001). doi: 10.1006/meth.2001.1262; pmid: 11846609
75. J. A. C. Mondragon *et al*., JASPAR TFBS LOLA databases - Part 1, Zenodo (2022). doi: 10.5281/zenodo.6860527

ACKNOWLEDGMENTS

We thank A. Killilea, C. Tasto, M. Fischer, and W. Hercule of the UC Berkeley Cell Culture Facility for cell culture support. We thank J. Marchand, M. Neugebauer, K. Thevasundaram, D. Miller, M. Sosa, and M. Chang for use of and technical assistance with their QqQ mass spectrometer. We thank Metabolon, K. Jackson, and P. Ramamoorthy for their work on performing the untargeted global metabolomics. We thank EpigenDx and A. Hopkins for their work on designing and performing the CpG methylation analysis. We thank M. L. Tomasi and S. C. Lu for providing their pGL3 basic luciferase plasmid with the *MAT1A* promoter. **Funding:** We thank the NIH (ES28096 and ES4705 to C.J.C. and D.K.N., and GM139245 to C.J.C.) for research support. K.J.B., V.N.P., and C.C.W. thank the National Science Foundation for graduate research fellowships, and C.C.W. thanks the NIH F31 for a graduate fellowship. V.N.P. and C.C.W. were partially supported by the NIH Chemical Biology Interface Training grant T32 GM066698. J.D.W.T. thanks A*STAR for a postdoctoral fellowship. F.A.D., C.L.M., and K.J.P. were funded by the Medical Research Council. F.A.D. was supported by CRUK (C42693/A23273), C.L.M. by the Wellcome Trust (217267/Z/19/Z), and K.J.P. by the Medical Research Council (MC_UU_00029/1). L.B.P. is supported by MCIN/AEI /10.13039/501100011033 and European Union NextGenerationEU/PRTR (Ayuda RYC2021-032395-I) and Departament de Recerca i Universitats / Generalitat de Catalunya (2021 SGR 01309). M.E. is supported by MCIN/AEI /10.13039/501100011033/ and the European Regional Development Fund, "A way to make Europe" ERDF (project PID2021-125282OB-I00) and Departament de Recerca i Universitats/Generalitat de Catalunya (2021 SGR 01494). M.C.P.-H. and N.T.I. were supported by the UC Cancer Research Coordinating Committee CRN-19-585450. For the purpose of Open Access, the authors have applied a CC BY copyright license to the Author Accepted Manuscript.
Author contributions: V.N.P., K.J.B., and C.J.C. conceived the study. V.N.P., K.J.B., C.C.W., D.K.N., and C.J.C. designed, performed, and analyzed the proteomics experiments. V.N.P., K.J.B., and C.J.C. designed, performed, and analyzed the formaldehyde-cysteine modification experiment, MAT biochemical assays, and cell-based experiments. J.D.W.T. and C.J.C. designed and performed experiments for MAT1A crystallography. E.J.G. performed additional MAT activity assays. L.T. performed formaldehyde fluorescence imaging. V.N.P. generated the *MAT2A* KO and *MAT1A* KO HepG2 cell lines with assistance from M.C.P.-H. for CRISPR-mediated knockout. V.N.P., F.A.D., C.L.M., and K.J.P. designed, performed, and analyzed all mouse experiments. V.N.P. and N.T.I. designed experiments and statistical analyses for MAT CpG methylation analysis. C.A.G.-P., L.B.P., and M.E. performed and analyzed genome-wide DNA methylation experiments. V.N.P. and C.J.C. designed and performed cellular knockdown and luciferase promoter assays. V.N.P., K.J.B., and C.J.C. wrote the manuscript with input from other authors. All authors reviewed the manuscript. **Competing interests:** D.K.N. is a cofounder, shareholder, and adviser for Artris Therapeutics. D.K.N. is a cofounder, shareholder, and scientific advisory board member for Frontier Medicines and Vicinitas Therapeutics. D.K.N. is a member of the board of directors for Vicinitas Therapeutics. D.K.N. is also on the scientific advisory board of The Mark Foundation for Cancer Research, Photys Therapeutics, Oerth Bio, and Apertor Pharmaceuticals. D.K.N. is also an Investment Advisory Partner for a16z Bio, an advisory board member for Droia Ventures, and an iPartner at The Column Group. K.J.P. is Chief Scientist for Cancer Research UK. All other authors declare no competing interests. **Data and materials availability:** The raw mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium through PRIDE with the dataset identifier PXD018251. The MAT1A crystal structure has been deposited to RCSB PDB as 8SWA. HepG2 *MAT1A* and *MAT2A* KO cell lines are available upon request. The raw DNA methylation data has been deposited to the Gene Expression Omnibus database repository with the dataset ID GSE242266 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE242266). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by the Wellcome Trust (217267/Z/19/Z), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abp9201
Supplementary Text
Materials and Methods
Figs. S1 to S14
Tables S1 to S7
References (*76*–*80*)
MDAR Reproducibility Checklist

Submitted 7 March 2022; resubmitted 21 June 2023
Accepted 24 September 2023
10.1126/science.abp9201

RESEARCH ARTICLE SUMMARY

GENE EXPRESSION

Sex-biased gene expression across mammalian organ development and evolution

Leticia Rodríguez-Montes*, Svetlana Ovchinnikova, Xuefei Yuan, Tania Studer, Ioannis Sarropoulos, Simon Anders, Henrik Kaessmann*†, Margarida Cardoso-Moreira*†

INTRODUCTION: Sexually dimorphic traits are widespread among mammals and emerge during development through sex-specific gene expression programs. We know very little about these programs, including the genes, regulatory networks, and cell types that underlie them. It is also unclear when differences between the sexes emerge during organ development and how sex differences evolve across species.

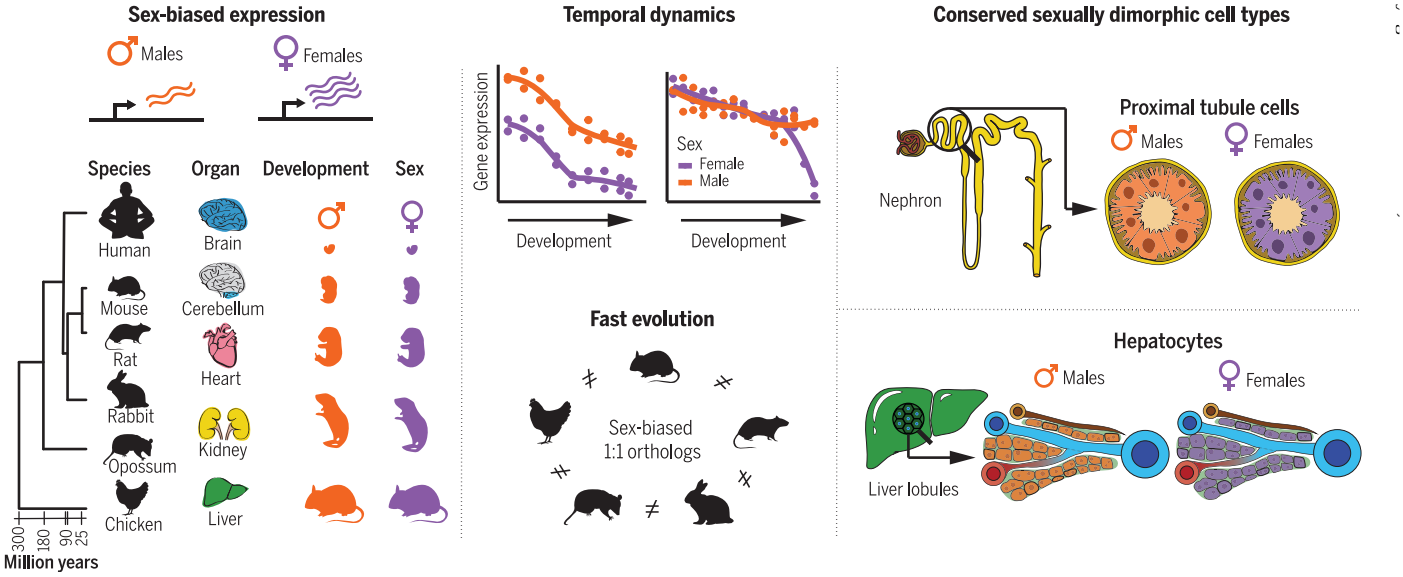
RATIONALE: To investigate the levels, temporal dynamics, and evolution of sex-biased expression during organ development, we analyzed RNA-sequencing (RNA-seq) time-series data from six species (human, mouse, rat, rabbit, opossum, and chicken) covering the development of five organs (brain, cerebellum, heart, kidney, and liver). Next, we determined the cell-type specificity of sex-biased expression and its conservation across species by analyzing new and existing single-cell RNA-seq datasets. Finally, to understand the mechanisms underlying sex differences in gene expression, we analyzed existing chromatin immunoprecipitation–sequencing datasets to identify the transcription factors responsible for the observed sex-specific gene expression programs.

RESULTS: We found that the levels of sex-biased expression varied substantially across organs and species. Across mammals, sex-biased expression was rare during organ development. In organs with high levels of sex-biased expression, the differences between the sexes only appeared around the time of sexual maturity. Sex-biased expression has evolved rapidly at the gene level, with only a few sex-biased genes conserved among placental mammals. The conserved gene set includes long noncoding RNAs involved in X chromosome inactivation and the X-linked gametologs of ubiquitously expressed Y genes. Unexpectedly, we found that the evolutionary mechanisms underlying this fast evolution differed between organs. In some organs, such as the rabbit heart or the mouse kidney, evolutionarily old genes have quickly evolved sex-biased expression through changes in gene regulation. In other organs, such as the mouse and rat liver, newly evolved genes have driven sex differences, with species differences arising through species-specific genes. Nevertheless, although we found these genes were species specific, they often belonged to gene families with sex-related functions and sex-biased members in multiple species, indicating cross-species conservation at the level of gene families. We also found that in organs with high levels of sex-biased expression, sex differences were often limited to specific cell types. These cell types were conserved across species in their sexual dimorphism. These results demonstrate that even though sex-biased expression evolves fast at the gene level, it evolves slowly at the level of cell types. Finally, we show that sex-biased expression results from a combinatorial process involving multiple transcription factors and can evolve quickly through the gain or loss of binding sites for a subset of these genes.

CONCLUSION: This study investigated sex differences in gene expression throughout the development of five organs in six species. We found that sex-biased expression varied substantially across organs and species and was often cell-type specific. Sex differences in gene expression were rare during organ development, increasing abruptly only around sexual maturity. Finally, sex-biased expression evolved quickly across species between orthologous genes. However, the cell types responsible for most sex-biased expression were the same across species. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: l.montes@zmbh.uni-heidelberg.de (L.R.-M.); h.kaessmann@zmbh.uni-heidelberg.de (H.K.); margarida.cardosomoreira@crick.ac.uk (M.C.-M.)
†These authors jointly supervised the work.
Cite this article as L. Rodríguez-Montes *et al.*, *Science* **382**, eadf1046 (2023). DOI: 10.1126/science.adf1046

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf1046



Sex-biased expression across the development of five organs in six species. Our analysis of a multispecies time-series gene expression dataset revealed the temporal dynamics of sex-biased expression during organ development and showed that it evolves quickly at the gene level but slowly at the cell-type level.

RESEARCH ARTICLE

GENE EXPRESSION

Sex-biased gene expression across mammalian organ development and evolution

Leticia Rodríguez-Montes[1]*, Svetlana Ovchinnikova[2], Xuefei Yuan[1], Tania Studer[1], Ioannis Sarropoulos[1], Simon Anders[2], Henrik Kaessmann[1]*†, Margarida Cardoso-Moreira[3]*†

Sexually dimorphic traits are common among mammals and are specified during development through the deployment of sex-specific genetic programs. Because little is known about these programs, we investigated them using a resource of gene expression profiles in males and females throughout the development of five organs in five mammals (human, mouse, rat, rabbit, and opossum) and a bird (chicken). We found that sex-biased gene expression varied considerably across organs and species and was often cell-type specific. Sex differences increased abruptly around sexual maturity instead of increasing gradually during organ development. Finally, sex-biased gene expression evolved rapidly at the gene level, with differences between organs in the evolutionary mechanisms used, but more slowly at the cellular level, with the same cell types being sexually dimorphic across species.

In many vertebrates, sex differences are the most extreme phenotypic variation seen within species (*1*). Although some sexually dimorphic traits are evident to the naked eye (e.g., differences in body size or plumage), many are not visible but are no less important (e.g., differences in drug clearance or immune responses) (*2*, *3*). Sexually dimorphic traits are specified at different points during development through sex-specific gene expression programs. Males and females are almost identical genetically, only differing in their sex chromosomes (X and Y in mammals, Z and W in birds). Genes on these chromosomes (e.g., *SRY* in mammals and *DMRT1* in birds) initiate the sex-determination pathways responsible for the differentiation of the gonads into ovary or testis (*1*, *4*). Upon sexual differentiation, the gonads start producing sex hormones (e.g., testosterone and estrogens) that reach different parts of the body and bind to their receptors on target cells. The hormonal signals trigger gene-regulatory cascades that differ between males and females, leading to differential gene expression between the sexes and the development of sexually dimorphic traits (*5*). In mammals, the development of most sexual phenotypes depends on sex hormones (*6*), whereas in birds, although sex hormones still play a role, sexual phenotypes are largely cell autonomous, with somatic cells carrying an inherent sex identity (*7*).

Genes with sexually dimorphic expression are called "sex-biased" genes and include those expressed exclusively in one sex and those expressed in both sexes but at different levels. Sex-biased genes are called male biased or female biased depending on which sex shows the higher expression level. Apart from the sex-determination pathways, little is known about sex-specific developmental gene expression programs. Most studies on sex-biased gene expression have focused on adults [e.g., (*8*–*10*)], when phenotypic sex differences are greatest (*11*). However, some sexually dimorphic phenotypes are established early in development (*12*, *13*), and it is unclear when the sexual dimorphisms observed in adults first emerge. Therefore, it is critical to study sex differences in a developmental context.

Across animals as diverse as butterflies, flies, and birds, there is a high species turnover of sex-biased expression (*14*–*16*). Although the same is likely true for mammals, the reported extent of conservation of sex-biased expression differs between studies (*8*, *17*–*19*). Understanding the evolution of sex-biased expression within a developmental framework is essential because evolutionary and developmental processes are strongly intertwined, with species differences being usually lower early in organ development (*20*, *21*). Here, we used bulk RNA sequencing (RNA-seq) time-series datasets to describe the extent, temporal dynamics, and conservation of sex-biased gene expression across organs and species. We then combined these data with single-cell RNA-seq (scRNA-seq) and chromatin immunoprecipitation–sequencing (ChIP-seq) datasets to infer the cellular and molecular mechanisms responsible for sexually dimorphic expression and their evolution across mammals.

[1]Center for Molecular Biology of Heidelberg University (ZMBH), DKFZ-ZMBH Alliance, D-69120 Heidelberg, Germany. [2]BioQuant, Heidelberg University, D-69120 Heidelberg, Germany. [3]Evolutionary Developmental Biology Laboratory, The Francis Crick Institute, London NW1 1AT, UK.
***Corresponding author. Email: l.montes@zmbh.uni-heidelberg.de (L.R.-M.); h.kaessmann@zmbh.uni-heidelberg.de (H.K.); margarida.cardosomoreira@crick.ac.uk (M.C.-M.)**
†These authors jointly supervised the work.

Extent of sex-biased expression across organs and species

To study sex-biased expression during organ development, we analyzed RNA-seq time-series data (*20*) from six species (human, mouse, rat, rabbit, opossum, and chicken) covering the development of five organs (brain, cerebellum, heart, kidney, and liver) (Fig. 1A). The time series spanned from early organogenesis to adulthood (7 to 16 stages) with one to three replicates per sex per stage, except for human, where the time series end shortly after birth, often with only one replicate per sex (table S1). We identified sex-biased genes using an approach that combined information from four time-series differential expression algorithms that we validated through extensive simulations (*22*). Because the human time series ended shortly after birth, we adopted a different strategy to call sex-biased expression in humans. We required the set of genes identified as sex biased during prenatal development to also be classified as sex biased in adults using the Genotype Tissue Expression (GTEx) resource (*9*) (table S7). Therefore, the set of human sex-biased genes was composed only of genes that differ in adults and started differing between the sexes pre- or perinatally. By contrast, the sets of sex-biased genes for the other species included genes that were differentially expressed at any point during development (*22*). The gene expression profiles and sex-bias status can be explored interactively at https://apps.kaessmannlab.org/sexbiasapp.

We first examined the levels of sex-biased expression in mouse, rat, rabbit, opossum, and chicken. We found considerable differences between species and organs in the levels of sex-biased expression. Chicken had the highest percentage of sex-biased genes, with 8% of the genes tested being sex biased in at least one organ (1337 genes). Across mammals, mouse had the highest percentage of sex-biased genes (2127, 5.9% of all genes tested), followed by rat (1005 genes, 3.9%), with rabbit and opossum showing considerably fewer sex-biased genes (287 and 200 genes, respectively, ~1%). The lower number of sex-biased genes in opossum is at least partly a consequence of the lower sensitivity of our approach in this species (fig. S1C) (*22*). Within each species, the number of sex-biased genes varied extensively across organs (Fig. 1B). For example, in mouse, only 15 genes were sex biased in the brain, whereas 1891 genes were sex biased in the kidney. The organs with the most sexually dimorphic transcriptomes differed between species: kidney in mouse and rat, heart in rabbit, liver in opossum, and brain in chicken (Fig. 1B).

Most genes were sex biased in a single organ, as previously observed (*8*–*10*) (Fig. 1C). However, these genes tended to be expressed in

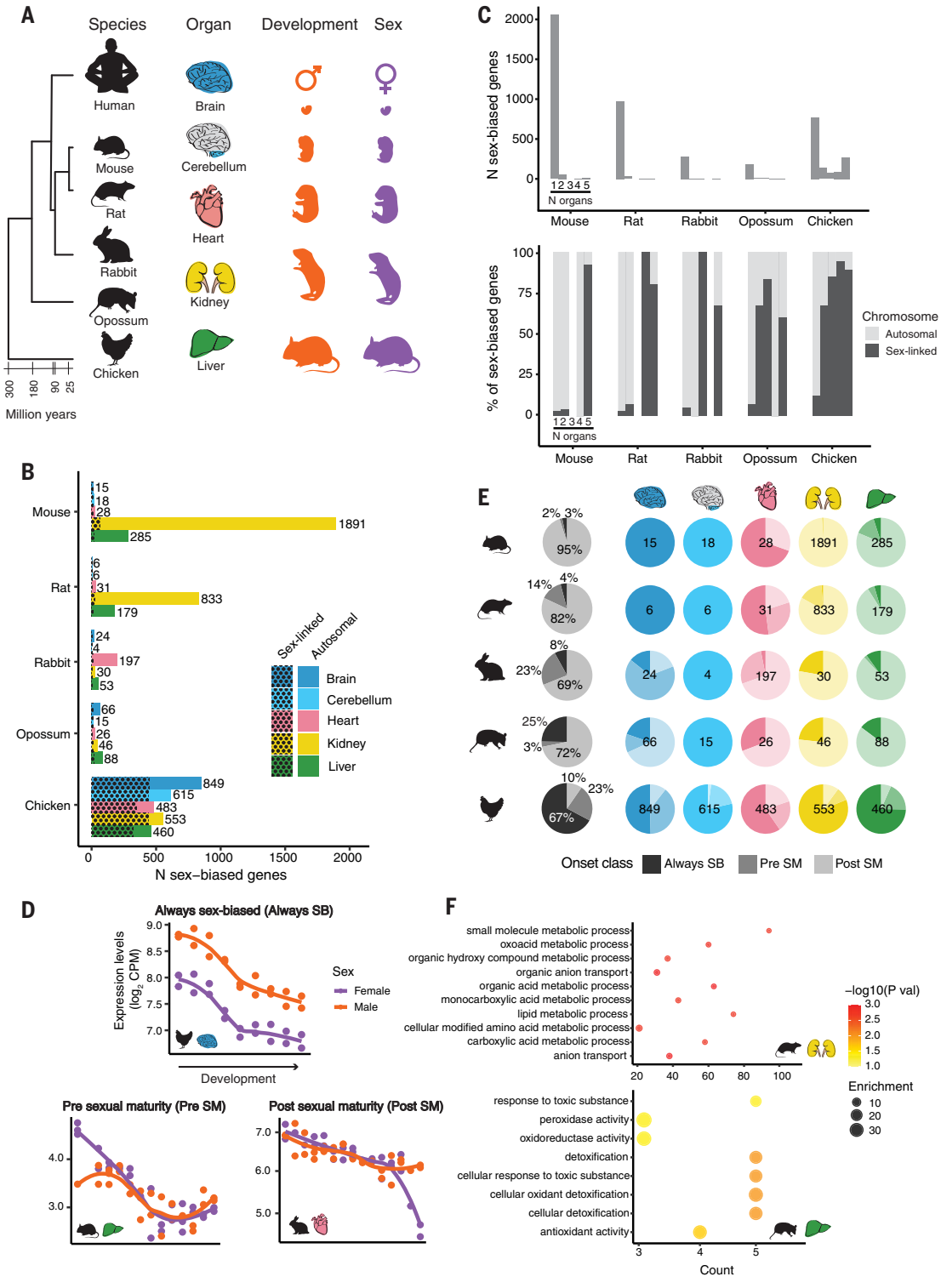


Fig. 1. Extent and onset of sex-biased gene expression. (**A**) Summary of the dataset. (**B**) Number of sex-biased genes by species and organ. Spotted pattern indicates genes located on sex chromosomes, X or Y in mammals (for rat, rabbit, and opossum, these only include X-linked genes because Y-linked genes were not present in the assemblies) and W or Z in chicken. (**C**) Number of sex-biased genes and chromosomal location as a function of the number of organs in which genes were sex biased. (**D**) Examples of genes belonging to each of the onset classes, *RPL17* in chicken brain, *Pagr1a* in mouse liver, and *LUC7L* in rabbit heart. CPM, counts per million. (**E**) Percentage of sex-biased genes belonging to each of the onset classes: always sex-biased (Always SB), sex-biased pre–sexual maturity (Pre SM), or sex-biased post–sexual maturity (Post SM). Depending on the species, 0.001 to 0.03% of genes were not assigned to any of the three categories and are not shown in the plot (*22*). Shown is the total number of sex-biased genes per organ and species inside each pie plot. (**F**) Enriched biological processes among genes that become sex biased after sexual maturity in rat kidney and opossum liver (n = 688 in rat and n = 75 in opossum; Benjamini-Hochberg adjusted P < 0.05, hypergeometric test).

multiple organs, with only a minority of genes that were sex biased in one organ (~3 to 9% across species) being specifically expressed in that organ [organ-specificity index (τ) > 0.8]. In mammals, only a few genes were sex biased across multiple organs, and these were strongly enriched for genes on the sex chromosomes. Although genes on sex chromosomes contributed to ~3 to 7% of organ-specific sex-biased genes (consistent with ~2 to 10% of genes in each species being on sex chromosomes), they made up ~60 to 90% of genes that were sex biased across all organs ($P < 0.01$ in all species, X^2 test; Fig. 1C). This latter category included Y-linked genes, long noncoding RNAs involved in X chromosome inactivation (e.g., *XIST* in placental mammals and *RSX* in opossum), X-linked genes that escaped X chromosome inactivation, and a small number of autosomal genes (e.g., *Uba5* in mouse) (tables S2 to S7) (*22*). In contrast to mammals, in chicken, hundreds of genes were sex biased across multiple organs. Most of these genes are on the Z chromosome and reflect the lack of a global dosage compensation mechanism (*23*). Because males have two Z chromosomes and females only one, most Z-linked genes show consistent higher expression in males than in females across multiple organs (461 genes; fig. S2). In all species, genes that are sex biased across multiple organs predominantly have the same direction of bias (male or female bias) across organs (from 78% of multiorgan sex-biased genes in rat to 100% in opossum).

Onset of sex-biased expression

Because sexually dimorphic phenotypes are most noticeable in adults, we expected adults to show the highest levels of sex-biased expression (*11*). However, it is unkown how much sex-biased expression exists during organ development and when the onset of the sex differences observed in adults occurs. To answer these questions, we determined the onset of sexually dimorphic expression for each sex-biased gene using soft clustering (*22*, *24*) (tables S8 to S13). We consistently found three classes of sex-biased genes: (i) genes sex biased across all developmental stages, (ii) genes sex biased before sexual maturity, and (iii) genes that become sex biased around or after sexual maturity (Fig. 1D).

In mammals, most genes became sex biased around or after sexual maturity (69 to 95% of sex-biased genes, depending on the species) (Fig. 1E). These genes were enriched among functions specific to each organ (Fig. 1F and fig. S3, B to D), including detoxification in the mouse, rat, and opossum liver and transport of small molecules in the mouse and rat kidney. These enrichments implicate sex-biased genes in the specific physiological processes executed by each organ and may underlie known sex differences in these processes (*25*–*27*).

A considerably smaller fraction of genes showed differences before sexual maturity (2 to 23% of all sex-biased genes) (Fig. 1E). Among these were genes that started to differ between the sexes before sexual maturity and continued to do so in adults (8 genes in mouse liver, 7 genes in rat heart, 14 genes in rabbit brain, and 23 genes in rabbit heart); genes that differed across several time points but were similarly expressed between the sexes in adults (8 genes in opossum brain and 130 genes in rat kidney); and genes that were sex biased only during early development (31 genes in mouse liver, 8 genes in rat liver, 31 genes in rabbit heart, and 1 gene in rabbit liver). Many of the mouse genes that were sex biased before sexual maturity were associated with sexually dimorphic phenotypes (table S14). The International Mouse Phenotyping Consortium (IMPC) (*28*) generated single-gene knockout lines for 10 of the 39 mouse genes with early onset of sex-biased expression, and five showed sexually dimorphic phenotypes (50 versus 14% of sexually dimorphic phenotypes among a total of 8619 knockouts, $P < 0.01$, X^2 test). Knockout of *Ndrg4*, a sex-biased gene in the liver, leads to increased levels of circulating creatinine and blood urea nitrogen in females but not in males (*28*). Similarly, males without a functional *Casq1* (sex biased in the liver) show abnormal cholesterol homeostasis, whereas females do not (*28*).

Finally, in mammals, some genes were sex biased across all developmental stages (3 to 25% of sex-biased genes) (Fig. 1E). These genes were also sex biased across multiple organs and were predominantly located on the sex chromosomes ($P < 0.01$, X^2 test). Therefore, there is a set of genes that are sex linked and sex biased throughout the entire development of multiple organs. This set includes Y-linked genes, long noncoding RNAs involved in X chromosome inactivation, and the small number of X gametologs of ubiquitously expressed Y-linked genes (e.g., *EIF2S3X*, *DDX3X*, *KDM6A*, and *KDM5C*), which escape X chromosome inactivation. There were also a few autosomal genes (fig. S3E) that were sex biased across the entire development of multiple organs, including *Uba5* in mouse, *Ddx3y* in rat (located on chromosome 13), a rabbit ortholog of the human *EIF1AY*, and *ZNF451*, and two more genes in opossum. Additionally, a few genes in each species were sex biased across all developmental stages in an organ-specific manner (e.g., *Vamp7* in the mouse heart and *Hip1r* in the rat kidney). Overall, we found that in organs with high levels of sex-biased expression (e.g., mouse liver), most genes became sex biased around or after sexual maturity, whereas in organs with low levels of sex-biased gene expression (e.g., mouse brain), most genes were always sex biased (Fig. 1E).

The set of human sex-biased genes consisted of genes that start differing between the sexes before or near birth and remain sex-biased in adults [as identified by GTEx (*22*)]. Among mammals, human had the largest number of genes in this category (78 genes), followed by rabbit (43 genes), mouse and rat (26 genes), and opossum (22 genes), although this analysis may be underpowered for the marsupial (fig. S3F) (*22*). These sex-biased genes were uniformly distributed among the organs, and whereas they were enriched for sex-chromosome genes, many were autosomal (fig. S3G).

In chicken, the temporal dynamics of sexually dimorphic expression were opposite to those in mammals (Fig. 1E). Only a minority of genes (~10% of all sex-biased genes) became sex biased around or after sexual maturation. Most genes (~67%) were sex biased across all developmental stages and across organs. Although most of these always sex-biased genes were sex linked (~85% of always sex-biased genes) and reflect the lack of global dosage compensation, 15% were autosomal.

Conservation of sex-biased expression

Next, we investigated the extent of conservation of sex-biased expression across species (i.e., mouse, rat, rabbit, opossum and chicken, with comparable time series). We determined the overlap between the sets of sex-biased genes in the different species according to the onset of sex-biased expression (i.e., always sex biased, sex biased before sexual maturity, or sex biased after sexual maturity). There were no sex-biased genes conserved between mammals and chicken or across all mammals (Fig. 2A and fig. S4A). We identified only five sex-biased genes conserved across mouse, rat, and rabbit (Y-linked genes excluded). These included three genes that were always sex biased (*Xist* and the X gametologs *Eif2s3x* and *Kdm6a*) and two genes that were sex biased in the liver after sexual maturity (*Cux2* and *Nipal1*). Except for *Nipal1*, these genes were also sex biased in the corresponding organs in adult humans (*9*), suggesting that they are conserved across placental mammals. For mouse, rat, and rabbit, there was little conservation outside the set of genes that were always sex biased (Fig. 2B and fig. S4B) (*22*).

Previous studies in adults (*8*, *17*–*19*) also found low levels of conservation of sex-biased expression, but the number of conserved sex-biased genes varied considerably between studies, ranging from a few (*17*, *18*) to several hundred (*8*, *19*). Because our approach for calling sex-biased genes is underpowered to detect sex differences limited to one or two stages (fig. S1) (*22*), which includes adult-only sex differences, we could be underestimating the extent of conservation of sex-biased expression. To test this possibility, we applied classical differential expression analysis using DESeq2 (*22*, *29*) to the adults in our dataset and independently to four prenatal stages. As expected, there was a good overlap between sex-biased

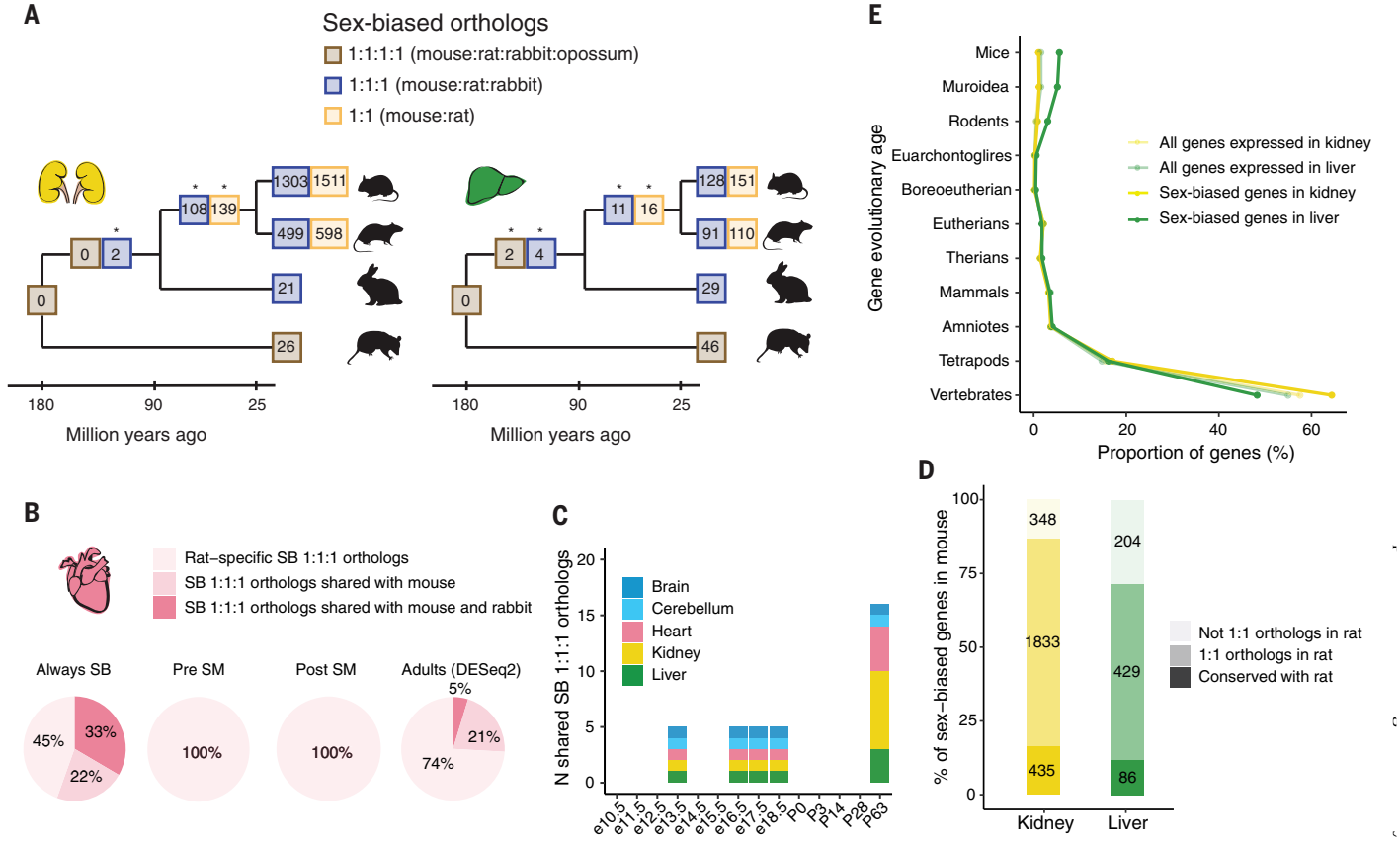


Fig. 2. Conservation of sex-biased gene expression. (**A**) Phylogeny showing the number of sex-biased orthologs in kidney and liver across mammals. *Benjamini-Hochberg adjusted $P < 0.05$, permutation test. The different numbers reflect the different sets of 1:1 orthologs used. For example, the set of 1:1 (mouse:rat) orthologs includes all 1:1:1 (mouse:rat:rabbit) orthologs plus genes that are only 1:1 orthologs between mouse and rat. (**B**) Percentage of sex-biased 1:1:1 orthologs in rat heart that were only sex biased in rat; sex biased in rat and mouse; or sex biased in rat, mouse, and rabbit, depending on the onset of sex-biased expression. (**C**) Number of shared sex-biased 1:1:1 orthologs in mouse, rat, and rabbit at different developmental stages (matched across species) using classical differential expression analysis (DESeq2). (**D**) Number and percentage of sex-biased genes in mouse kidney and liver that were also sex biased in rat, had a 1:1 ortholog in rat, or did not have a 1:1 ortholog in rat. (**E**) Proportion of expressed and sex-biased genes in the mouse kidney and liver and according to their evolutionary age.

genes identified in adults with DESeq2 and those identified by our time-series approach (fig. S4C) (*22*). The newly identified adult sex-biased genes showed significantly smaller differences in expression levels between the sexes than those identified by our time-series approach ($P < 0.0001$, Wilcoxon rank-sum test; fig. S4D). Using the set of adult sex-biased genes identified by DESeq2, we identified a higher number and proportion of sex-biased genes conserved across mouse, rat, and rabbit (17 versus five using the time-series approach; Fig. 2C and fig. S4B). By contrast, analysis of the four prenatal stages identified only *Xist* as being conserved across all three species (Fig. 2C). These results confirm that there were few sex-biased genes before sexual maturity, that only a few of them were conserved across species, and that these were better identified with our time-series approach. However, they also suggest that we were underestimating the extent of sex-biased expression in adults for genes with smaller differences in expression levels between the sexes. To overcome this limitation, we created an extended set of sex-biased genes that combined the time-series calls with those made with DESeq2 in adults.

Our analyses indicate that sexually dimorphic expression evolves rapidly, with conservation of sex-biased expression during development restricted to a few key genes, most of which are sex biased across all developmental stages and organs. These genes include *Xist*, the small number of X gametologs with partners on the Y chromosome that were ubiquitously expressed, and a few others (table S15).

Evolutionary age of sex-biased genes in mouse and rat

To explore the fast evolution of sex-biased expression, we focused on the two most closely related species, mouse and rat, using the extended set of sex-biased genes (combining the time-series and adult-only calls). In both species, the two most sexually dimorphic organs were the kidney and liver, but only a small percentage of genes were sex biased in both species (17% in kidney, 12% in liver; Fig. 2D). The conserved sex-biased genes tended to have the same direction of sex bias in the two species (64% in kidney and 73% in liver) and were involved in important processes, including transmembrane transport in the kidney and redox reactions in the liver.

Among the genes that were sex biased in mouse but not in rat, it is important to distinguish between those that had 1:1 orthologs in rat and those that did not (because of gene duplication or loss). Although in the kidney, most mouse sex-biased genes had 1:1 orthologs in the rat (87%), in the liver, 28% of mouse sex-biased genes did not. This suggests that many sex-biased genes in the mouse liver duplicated in the mouse and/or rat lineages or were lost in rat. To further explore these possibilities, we determined when mouse sex-biased genes first appeared during evolution, that is, when they first arose through gene duplication (*30*). For most genes in the kidney, the sex differences

were recent, but the genes themselves were old (Fig. 2E). By contrast, in the liver, many sex-biased genes appeared recently in evolution, with at least 5% being mouse specific (a likely underestimation because we could not assign an evolutionary age for 10% of sex-biased genes in the liver versus only 5% in the kidney) (*22*).

Because newly emerged genes are among the least studied (*31*, *32*), we manually examined the annotations for these genes and found that many derived from the expansion of three gene families, the cytochrome P450 family (20 of 77 genes), the major urinary protein family (16 of 77 genes), and the Slc22 transporter family (5 of 77 genes). All three families have undergone successive gene duplications (*33*–*35*) and are involved in critical sex functions. The cytochrome P450 family is involved in the metabolism of xenobiotics and the transformation of endobiotics such as steroid hormones, processes with many sex differences (*27*). The major urinary protein family codes for sex pheromones and is involved in creating scent marks used for male-male competition, female assessment of males, and kin recognition (*33*). The rodent-specific expansion of the Slc22 transporter family has been associated with the transport of conjugated sex hormones (*35*). In the rat liver, recently emerged sex-biased genes also belong to the major urinary protein family (4 of 37 genes), which has expanded in parallel in mouse and rat (*36*, *37*), and the P450 family (2 of 37 genes). Although there are few conserved sex-biased orthologs in the mouse and rat liver, there is conservation of sex-biased expression at the level of gene families.

Our analysis of the evolutionary age of sex-biased genes uncovered important differences between organs. In the mouse and rat kidney, evolutionarily old genes quickly evolved sex differences in expression. This was also true for the rabbit heart (fig. S4E), the most sexually dimorphic organ in this species, where most sex-biased genes have 1:1 orthologs in mouse. However, in the liver, the evolution of sex differences often involved the evolution of new genes through the independent expansion of the same gene families.

Cellular basis of sex-biased expression

Sex differences in expression at the bulk tissue level can be a consequence of the differential expression of genes in the same cell types between males and females, differences in the proportion of cell types between the sexes, and/or genes being expressed in different cell types between males and females. Distinguishing between these scenarios requires investigating sex differences at the single-cell level. Therefore, we used single-cell datasets to identify the cell population(s) that expressed the sets of sex-biased genes identified at the bulk tissue level (*22*).

We focused on the two most sexually dimorphic organs in mouse, the kidney and liver. We generated a single-nucleus RNA-seq (snRNA-seq) dataset for four adult mouse liver samples (*22*) and used an existing scRNA-seq dataset for the adult mouse kidney (*38*). We then determined the expression of the extended set of sex-biased genes (combining the time-series and adult-only calls) in the single-cell datasets (*22*). In the mouse liver, male- and female-biased genes were specifically expressed in hepatocytes (Fig. 3, A and B), as previously observed (*39*). Male-biased genes were more highly expressed in male hepatocytes, and female-biased genes were more highly expressed in female hepatocytes (Fig. 3C). By contrast, in the mouse kidney, the cell-type specificity of sex-biased genes differed between male-biased and female-biased genes. Male-biased genes were expressed specifically in the proximal tubule cells, whereas female-biased genes were not cell-type specific and were expressed across several cell types in addition to the proximal tubule cells (Fig. 4, A and B). Despite this difference in cell-type specificity, the expression differences between the sexes were mostly restricted to the proximal tubule cells, as previously suggested (*38*). In this cell population, male cells expressed male-biased genes at higher levels than female cells, which in turn expressed more highly female-biased genes than male cells (Fig. 4C).

Studies suggested small morphological differences between males and females in these two organs, which could reflect cell composition differences (*40*–*42*). However, there is no evidence yet for differences in cell-type proportions between the sexes, and because of the levels of technical variability in the single-cell datasets, we cannot reliably address this question with these data (fig. S5A). Although we cannot discard the possibility that there are also differences in the abundance of proximal tubule cells in the kidney and hepatocytes in the liver between males and females, our data support that most sex differences in these two organs are the result of there being a female version and a male version of these two cell types and thus that male- and female-biased genes are differentially expressed between the sexes in the same cell types.

In the mouse kidney and liver, most sex-biased genes were only sex-biased starting around puberty (Fig. 1E), so we did not expect sexually dimorphic expression of these genes before birth. However, we wanted to know where these genes were expressed before showing sex differences. Therefore, we reanalyzed prenatal scRNA-seq datasets from the mouse kidney and liver (*43*, *44*). In the kidney, male-biased genes were already specifically expressed in the proximal tubule cells prenatally, whereas female-biased genes were more broadly expressed (as observed in adults). As expected, before birth, male and female cells expressed sex-biased genes at similar levels (or with considerably smaller differences than those observed in adults) (Fig. 4, D to F). A similar pattern was observed in the prenatal mouse liver. Before birth, male- and female-biased genes were expressed specifically in hepatocytes but with only minor differences in gene expression between male and female cells (Fig. 3, D to F). These results show that sex-biased genes are expressed in the same cell types prenatally as in adults and that only after puberty do the cells from males and females start to diverge in their expression.

Conservation of the sexually dimorphic cell types

Next, we investigated whether the cell-type specificity of sex-biased expression is conserved across species and, if so, if sex-biased expression is associated with the same cell types across species. To this end, we reanalyzed a scRNA-seq dataset for the rat kidney and liver (*45*). When we assessed the cell-type specificity of the extended set of rat sex-biased genes in the two organs, we identified the same two cell types that we had found in mouse. Male-biased genes were specific to the proximal-tubule cells in the kidney (Fig. 4, G to I), and male- and female-biased genes were specific to hepatocytes in the liver (Fig. 3, G to I). In both organs, the expression differences between male and female cells were the same as in mouse. These observations were not driven by the set of sex-biased genes common to both species (fig. S5, B and C). Our results suggest that although sex-biased expression evolves fast at the gene level, it evolves more slowly at the cell-type level. This is consistent with a single-cell study of the human kidney (*46*) that also identified the proximal tubule cells as being sexually dimorphic and driving most sex dimorphisms despite there being very little conservation in the set of sex-biased genes between rodents and humans.

To understand how sex-biased expression evolved so quickly between mouse and rat, we focused on genes that were sex biased in only one of the species and determined their expression in the other (where they were not sex biased). We investigated whether these genes were expressed in the same cell types in mouse and rat despite the difference in their sex-biased status. In the kidney, we found this to be true. Rat-only male-biased genes were also specifically expressed in the proximal tubule cells in the mouse dataset and vice versa (fig. S5, D and E). In the liver, although mouse-only sex-biased genes were also specifically expressed in hepatocytes in rat, rat-only sex-biased genes were not as hepatocyte specific as in mouse (fig. S5, F and G). These results indicate a difference in both the sex- and the cell-type-specific regulation of these genes between species.

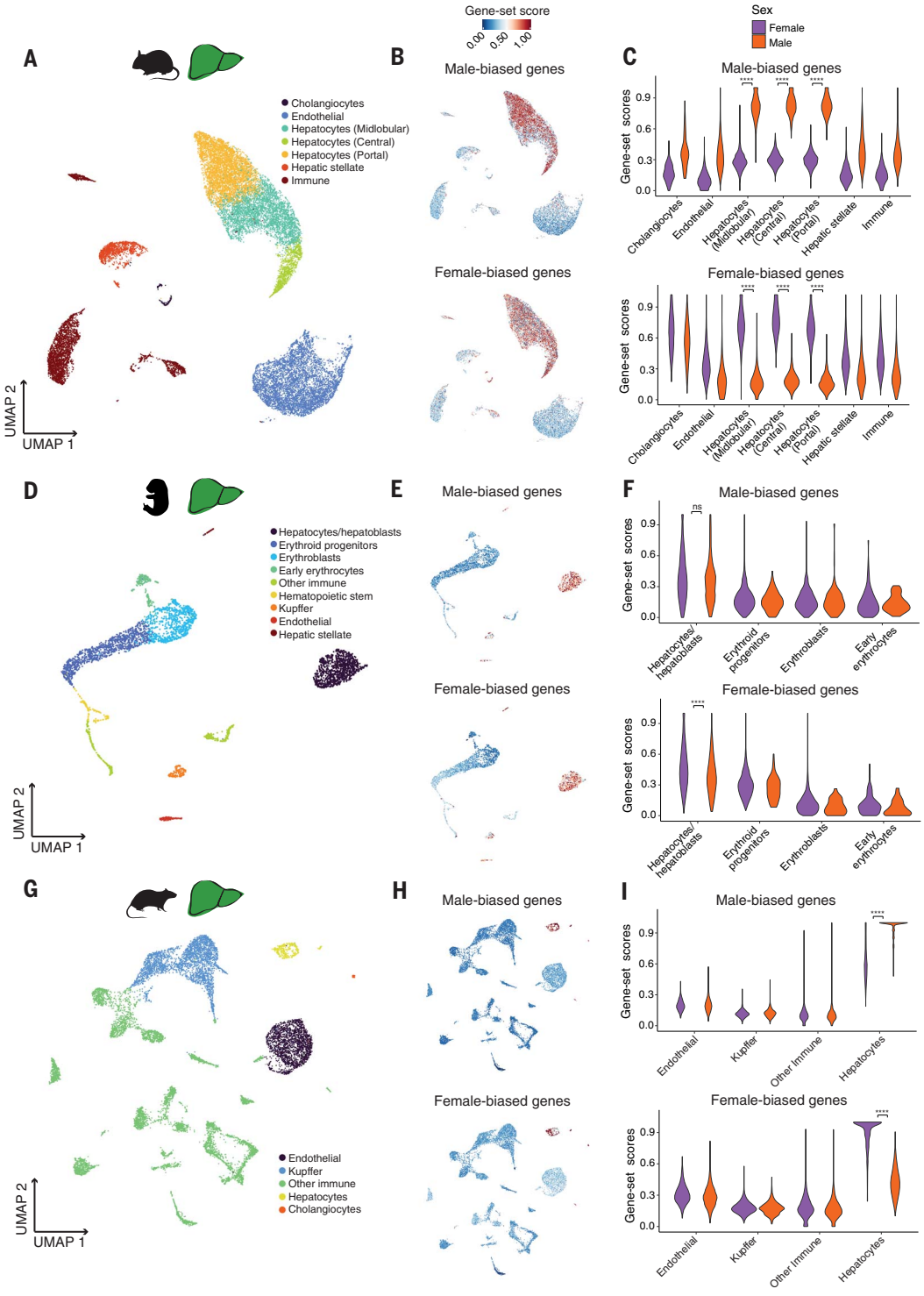


Fig. 3. Cellular basis of sex-biased gene expression in mouse and rat liver. (**A**) Uniform manifold approximation and projection (UMAP) of the adult mouse liver snRNA-seq dataset (22,512 cells). (**B**) UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in adult mouse liver. (**C**) Distribution of male-bias (up) and female-bias (down) gene-set scores according to cell type and separated by male and female cells in adult mouse liver (****Benjamini-Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test). (**D**) UMAP of prenatal mouse liver scRNA-seq dataset [data from (*44*)] (3847 cells). (**E**) UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in prenatal mouse liver. (**F**) Distribution of male-bias (up) and female-bias (down) gene-set scores according to cell type and separated by male and female cells in prenatal mouse liver (****Benjamini-Hochberg adjusted $P < 0.0001$; ns, nonsignificant, two-sided Wilcoxon rank-sum test). (**G**) UMAP of adult rat liver scRNA-seq dataset [data from (*45*)] (11,343 cells). (**H**) UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in the adult rat liver. (**I**) Distribution of male-bias (up) and female-bias (down) gene-set scores according to cell type and separated by male and female cells in adult rat liver (****Benjamini-Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test).

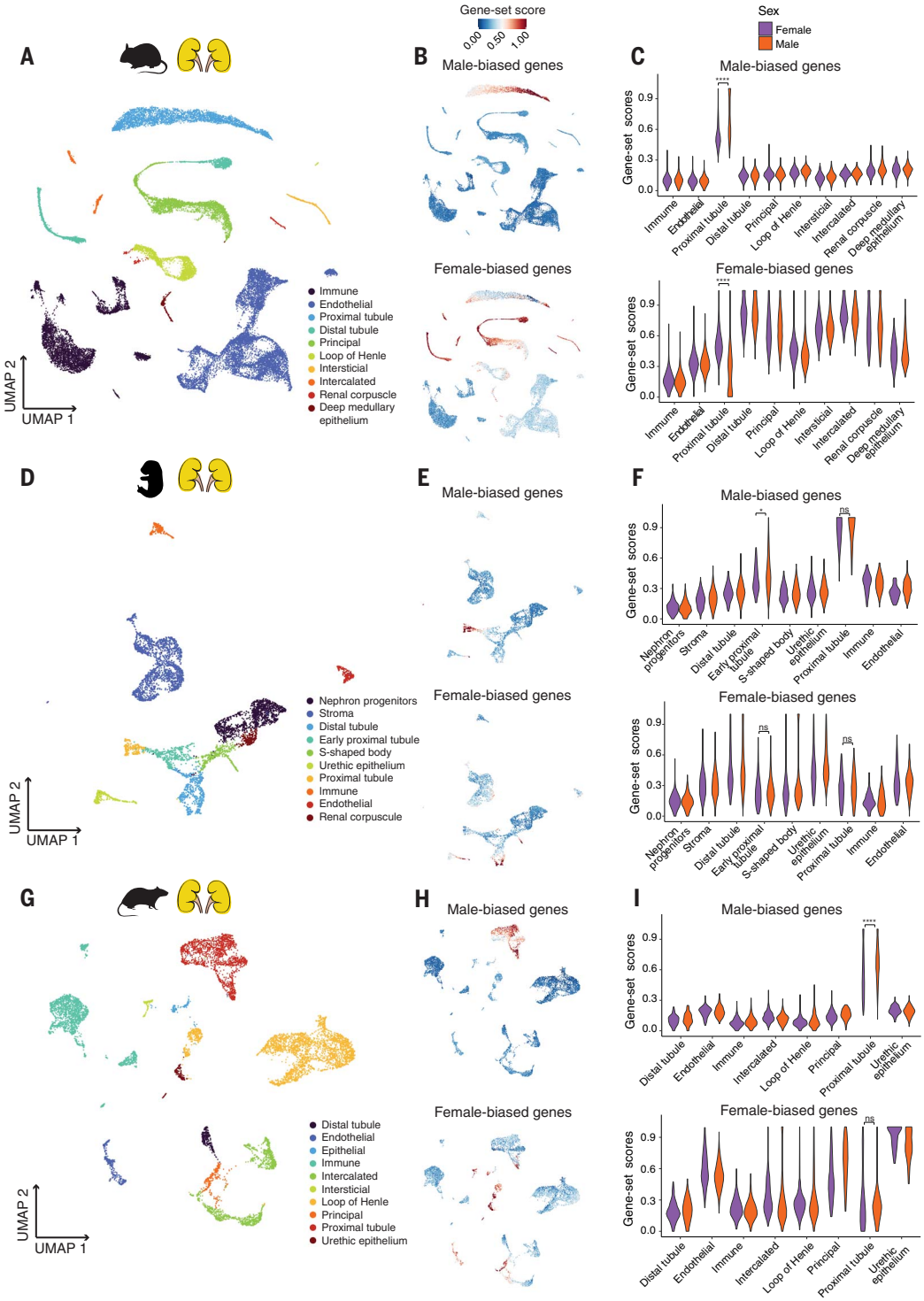


Fig. 4. Cellular basis of sex-biased gene expression in mouse and rat kidney. (**A**) UMAP of adult mouse kidney scRNA-seq dataset [data from (*38*)] (29,611 cells). (**B**) UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in adult mouse kidney. (**C**) Distribution of male-biased (up) and female-biased (down) gene-set scores according to cell type and separated by male and female cells in adult mouse kidney (****Benjamini-Hochberg adjusted $P < 0.0001$, two-sided Wilcoxon rank-sum test). (**D**) UMAP of prenatal mouse kidney scRNA-seq dataset [data from (*43*)] (5168 cells). (**E**) UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in prenatal mouse kidney. (**F**) Distribution of male-biased (up) and female-biased (down) gene-set scores according to cell type and separated by male and female cells in prenatal mouse kidney (*Benjamini-Hochberg adjusted $P < 0.05$; ns, nonsignificant, two-sided Wilcoxon rank-sum test). (**G**) UMAP of adult rat kidney scRNA-seq dataset [data from (*45*)] (9340 cells). (**H**) UMAPs illustrating expression of male-biased (up) and female-biased (down) genes in adult rat kidney. (**I**) Distribution of male-biased (up) and female-biased (down) gene-set scores according to cell type and separated by male and female cells in adult rat kidney (****Benjamini-Hochberg adjusted $P < 0.0001$; ns, nonsignificant, two-sided Wilcoxon rank-sum test).

Molecular basis of developmental sex-biased expression

Hormones play a critical role in sex-biased expression by differentially activating transcription factors (TFs) and their downstream targets in each sex (*47*). Although sex-biased TFs underlie a large fraction of sex-biased genes (*19*, *48*), sex-biased expression can also be achieved through non–sex-biased TFs, when, for example, differences in hormone concentrations lead to differential rates of TF translocation to the nucleus and transcriptional activation of downstream targets between the sexes (*49*). To identify the TFs responsible for the sex-biased expression in the mouse kidney and liver, we analyzed available ChIP-seq datasets for TFs responsive to the growth hormone [a key driver of sex differences in the liver (*49*, *50*)], TFs responsive to sex-related hormones (androgens and estrogens), and TFs that we classified here as being sex biased (*22*).

In the mouse kidney, male-biased genes were enriched among the targets of the androgen receptor (*Ar*) and a male-biased TF, *Hnf4a*, which is known to interact with *Ar* (Fig. 5A). *Ar* and *Hnf4a* were specifically expressed in the proximal tubule cells (Fig. 5C). By contrast, female-biased genes were enriched among the targets of *Ap-2*, a female-biased TF. In the kidney, 65% of sex-biased genes were targeted by at least one of these three TFs (compared with 20% of all genes expressed in the kidney, $P < 0.01$, X^2 test; Fig. 5B).

In the mouse liver, sex differences are mainly driven by the different temporal secretion patterns of the pituitary growth hormone, which is secreted continuously in females and in regular pulses in males (*50*–*53*). Both male- and female-biased genes in the mouse liver are enriched for known growth hormone–related TF targets, including *Stat5b*, a non–sex-biased TF (*54*); *Bcl6*, a male-biased transcriptional repressor (*55*, *56*); *Cux2*, a female-biased repressor (*57*); and *Hnf6*, another non–sex-biased TF (*58*) (Fig. 5A). TFs responsive to hormones can act as both inducers and repressors of gene expression depending on their interaction partners (*59*–*62*). Previous work showed that only 24% of *Stat5b*-binding sites were differentially bound by *Stat5b* in males and females (*55*). Accordingly, we found that male-biased genes were enriched for targets of male-enriched *Stat5b*-binding sites, whereas female-biased genes were enriched for targets of female-enriched *Stat5b*-binding sites (Fig. 5A). At the single-cell level, all of these TFs except for *Stat5b* were hepatocyte specific (Fig. 5C), as previously reported (*39*), and so were their targets (including those of *Stat5b*).

In the liver, male-biased genes were also enriched among the targets of sex hormone–responsive TFs, namely *Esr1* and *Ar* (Fig. 5A), which supports that sex hormones drive sex differences in the liver in addition to the growth hormone (*63*). However, whereas many sex-biased genes were targeted by both sets of hormones (Fig. 5E), only a few were targeted exclusively by sex hormone–responsive TFs, supporting a more prominent role in sexual dimorphism for the growth hormone.

To investigate the molecular basis of the fast evolution of sex-biased expression between mouse and rat, we looked at the distribution of binding sites for the TFs driving sex differences in the kidney and liver for genes that were sex biased only in rat (i.e., we examined their mouse orthologs). In the kidney, the mouse orthologs of rat-only female-biased genes were not enriched among the targets of *Ap-2*, in agreement with their non–sex-biased status (fig. S6A). However, the mouse orthologs of rat-only male-biased genes were still enriched among the targets of *Ar* and *Hnf4a* despite not being sex biased. The lack of male-biased expression in the mouse genes was confirmed by the single-cell datasets and so is unlikely to be a false negative (fig. S6, C and D). This result suggests that another element (e.g., another TF) is also necessary for male-biased expression and that it is its absence that drives the species difference. Alternatively, there could be quantitative differences in the binding of *Ar* and *Hnf4a* that explain the species difference.

In the liver, the mouse orthologs of rat-only female-biased genes differed from mouse female-biased genes by not being enriched among the targets of *Cux2* and the female-biased *Stat5b*-binding sites (fig. S6A), the two TFs that sit atop of the cascade leading to female-biased expression. The mouse orthologs of rat-only male-biased genes differed from mouse male-biased genes by not being enriched among the targets of *Cux2* and *Hnf6*, which work downstream of the growth hormone, and by not being enriched among the targets of sex hormones (i.e., *Esr1* and *Ar*). The cross-species comparisons for the liver and kidney support a model in which sex-biased expression is the result of a combinatorial process involving multiple TFs, and that sex-biased expression can evolve quickly through the gain or loss of binding sites for a subset of the intervening TFs.

TF binding is associated with chromatin accessibility. Previous studies identified chromatin regions with different accessibility between the livers of male and female mice (*64*), which are related to different abundances and distributions of epigenetic marks between the sexes (*65*). We reexamined these datasets and found that sex-biased DNase-hypersensitive sites (DHS) were associated with sex-biased genes (Fig. 5D). We also found that the active chromatin marks H3K4me1, H3K4me3, H3K27ac, and H3K36me3 were associated with male-biased genes in males and female-biased genes in females, whereas the repressive mark H3K27me3 was associated with female-biased genes in males (Fig. 5D). In total, 81% of sex-biased genes in the mouse liver were targeted by at least one of the growth hormone–related TFs, sex-biased DHS, or sex-biased chromatin marks (compared with 58% of all genes expressed in the liver; $P < 0.01$, X^2 test; Fig. 5E).

The repressive mark H3K27me3 is introduced by the histone-modifying enzymes EZH1/2 (*66*) and removed by KDM6B and the gametologs UTY (male biased) and KDM6A (female biased) (Fig. 5F). Similarly, the activating mark H3K4me3 is introduced by SETD1a/SETD1b, MLL1/MLL2, and PRDM9 (*66*) and removed by KDM5A, KDM5B, and the gametologs KDM5D (male biased) and KDM5C (female biased). These gametolog pairs are among the rare genes that show conserved sex-biased expression across placental mammals. Perhaps despite the poor conservation of sex-biased genes across these species, sex differences in expression could ultimately involve similar molecular processes across placental mammals.

Discussion

We found that sex-biased expression varies substantially across species, organs, and developmental stages and that it is often cell-type specific. In mammals, sex-biased expression is rare during organ development. In sexually dimorphic organs, sex-biased expression abruptly increases around sexual maturity. We expected to find a large increase in the number of sex-biased genes with the onset of sexual maturity, as observed across species as diverse as frogs (*67*), stick insects (*68*), and human (*69*). However, we were surprised by the low levels of sex-biased expression during the development of organs with strong sex differences in adults. This suggests that in mammals, most sex differences only start at sexual maturity, when they are most visible. However, it is possible that some sex differences start before sexual maturity but are not reflected in sex-biased expression. There is evidence for this in our work, for example, in the TFs identified as driving sex-biased expression that were not themselves sex biased. In marked contrast to mammals, in chicken, most genes were sex biased across all developmental stages and most were Z linked. The lack of a global transcriptional dosage compensation mechanism on the Z chromosome means that ~5% of chicken genes are always differentially expressed between the sexes irrespective of the organ or developmental stage. This high level of sex-biased expression could underlie the dominance of cell-autonomous processes in driving sexual dimorphisms in birds.

In all species, a small set of sex-linked genes are sex biased throughout the development of multiple organs. This set contains most of the few genes with conserved sex-biased expression across placental mammals and includes the long noncoding RNAs involved in X chromosome

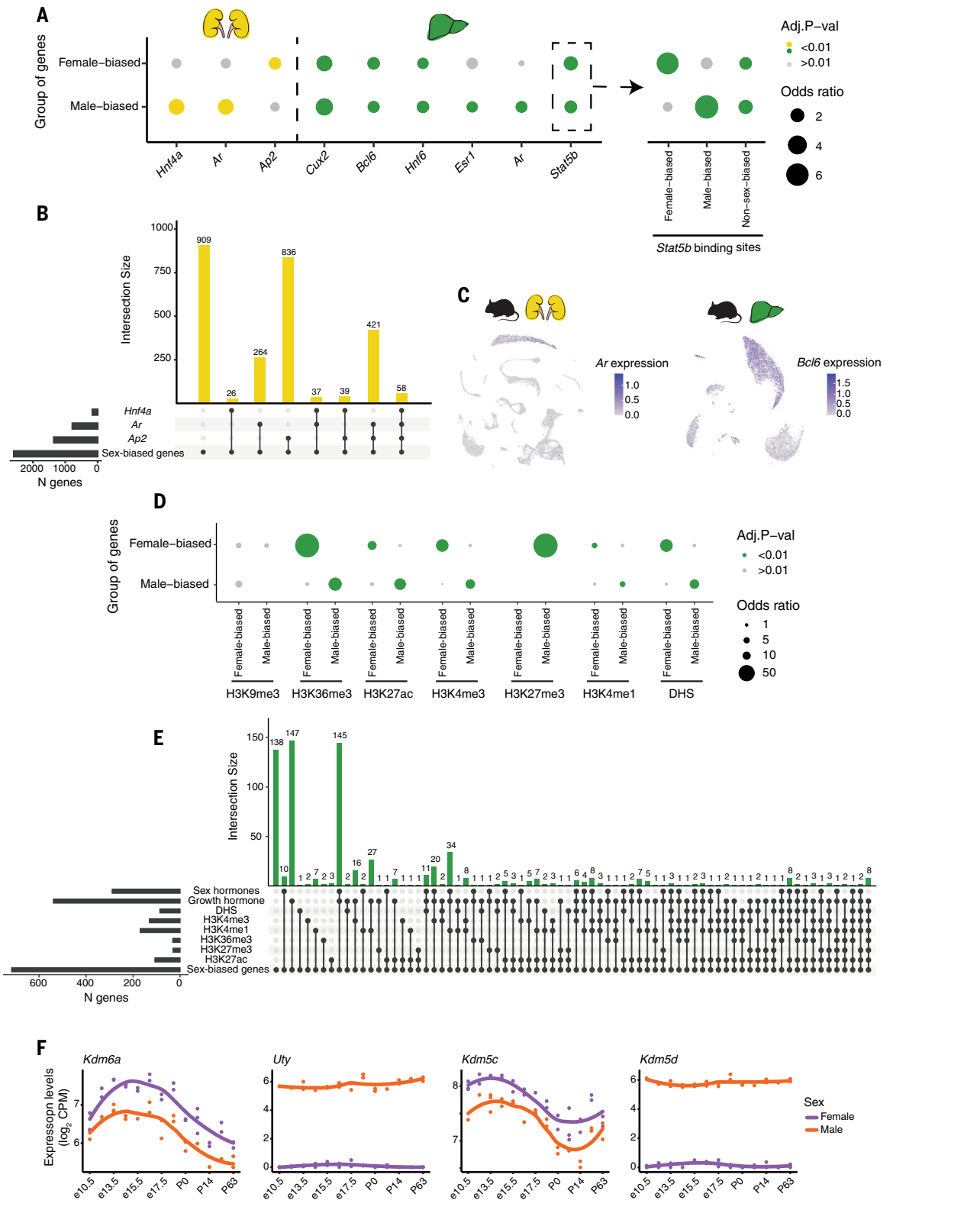


Fig. 5. Molecular basis of sex-biased gene expression. (**A**) Enrichment of sex-biased genes for genes regulated by hormone-responsive or sex-biased TFs in mouse kidney and liver. (**B**) Number of sex-biased genes targeted by hormone-responsive or sex-biased TFs in mouse kidney. (**C**) Examples of cell type–specific expression of two TFs (*Ar* in proximal tubule cells and *Bcl6* in hepatocytes). (**D**) Enrichment of sex-biased genes for genes located in close proximity to regions with sex differential distribution of epigenetic marks or sex-biased DHS. (**E**) Number of sex-biased genes targeted by hormone-responsive or sex-biased TFs or in close proximity to regions with sex differential distribution of epigenetic marks or sex-biased DHS sites in mouse liver. (**F**) Gene expression time courses of *Kdm6a/Uty* and *Kdm5c/Kdm5d* gametologs in mouse liver.

inactivation, ubiquitously expressed Y genes, and their X gametologs. In opossum, this set of sex-linked and always sex-biased genes is of special interest because they are prime candidates for underlying sex differences that occur before the differentiation of the gonads. Unlike in placental mammals, in marsupials, the development of some secondary sexual traits, such as the mammary gland and the scrotum, is independent of hormones (*70*). The development of these sexual traits depends instead on the number of X chromosomes (*71*), presumably a dosage difference in a hitherto-unknown X-linked gene (*72*, *73*). The small set of X-linked genes that we found to be consistently sex biased across organs and developmental stages are prime candidates (listed in table S16). These genes were not sex biased in placental mammals, and most started showing sex differences very early in opossum development, before the differentiation of the bipotential gonad (fig. S7). Two genes, *PHF6* and *DKC1*, are especially promising because, in human, they have been implicated in the development of the urogenital tract (*74*, *75*).

Work across various taxa suggests that sex-biased expression evolves quickly across species (*8*, *14*–*16*). Our work strongly supports this observation. Genes that are sex biased in one species are typically not sex biased in another, even among closely related species. However, our work shows that behind this general observation of fast evolution, there are key differences in the evolutionary mechanisms used among organs. In the rabbit heart or the mouse and rat kidney, evolutionarily old genes quickly evolved sex-biased expression through gains of sex-specific regulatory sequences. However, in the mouse and rat liver, newly evolved genes were the drivers of sex differences, with species differences arising through the deployment of species-specific genes. However, although these genes were species specific, they are members of gene families with sexually dimorphic members in multiple species. This suggests more conservation across species at the gene family level.

The rapid turnover of sex-biased expression across species could result from nonadaptive genetic drift or changes in patterns of natural or sexual selection. Although our study was not designed to estimate the contributions of these factors directly, some key results implicate changes in natural and sexual selection as drivers of species differences. First, although we observed rapid changes in the identity of the genes that were sex biased across species, we found sex differences to be limited to specific cell types and those cell types to be conserved across species in their sexual dimorphism. Genetic drift cannot easily account for these observations. Rather, it is more likely that some cells are a hotspot for sexually antagonistic traits because of their functions (e.g., the uptake and secretion of drugs and xenobiotics that affect the sexes differently). Natural selection can resolve intralocus sexual conflicts, which occur when genes have different expression optima in males and females (*76*, *77*), through the evolution of sex-biased expression. Second, several sex-biased genes, particularly in the liver, are known to mediate reproductive competition, which is highly suggestive of sexual selection (*14*). Several sex-biased genes in the liver belong to gene families involved in reproduction and mate choice, including the major urinary proteins, which encode for pheromones, and the cytochrome P450 family.

The exception to the fast evolution of sex-biased expression is the small number of genes in placental mammals that are always sex biased. Despite their small number, these conserved sex-biased genes could be playing important sex-related roles across species, as seems to be the case for the pairs of gametologs *KDM6A/UTY* and *KDM5C/KDM5D*. These pairs code for demethylases responsible for the removal of epigenetic marks that have different distributions in males and females and are involved in the regulation of expression of sex-biased genes in mouse and human (*9*, *65*, *78*–*81*). This suggests that in placental mammals, the genes that consistently show differences between the sexes during development could be involved in triggering and/or maintaining sex-specific developmental programs in each sex similarly across species.

Materials and methods summary

Detailed information on materials and methods is available as supplementary materials (*22*). In brief, we used four time-series differential expression algorithms to identify sex-biased genes across organ development: splineTimeR (*82*), DESeq2 (*29*), MaSigPro (*83*), and our own algorithm (*22*). The sets of sex-biased genes comprised genes called sex biased by at least two different time-series algorithms because this approach yielded the best results in terms of sensitivity and specificity, as shown by our extensive simulations [performed with seqgendiff (*22*, *84*)]. We used soft clustering as implemented in GPClust (*24*) to cluster genes in each organ and species according to their temporal profiles and to determine the onset of sex-biased expression.

We generated the mouse liver snRNA-seq dataset from four snap-frozen liver samples that were used for nuclei isolation and single-cell library construction using the Chromium Single Cell RNA Reagent kits (10x Genomics). We processed the sequencing data using Cell Ranger (*85*). Publicly available mouse and rat scRNA-seq datasets were obtained from the respective studies (*38*, *43*–*45*). All single-cell datasets were analyzed with Seurat (*86*), including quality control, dimensionality reduction, clustering, and cell-type annotation. We used ChIPseeker (*87*) to study TFs and epigenetic marks associated with sex-biased expression in publicly available datasets (*48*, *55*, *58*, *64*, *65*, *88*–*90*).

REFERENCES AND NOTES

1. J. Parsch, H. Ellegren, The evolutionary causes and consequences of sex-biased gene expression. *Nat. Rev. Genet.* **14**, 83–87 (2013). doi: 10.1038/nrg3376; pmid: 23329110
2. O. P. Soldin, D. R. Mattison, Sex differences in pharmacokinetics and pharmacodynamics. *Clin. Pharmacokinet.* **48**, 143–157 (2009). doi: 10.2165/00003088-200948030-00001; pmid: 19385708
3. S. L. Klein, K. L. Flanagan, Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016). doi: 10.1038/nri.2016.90; pmid: 27546235
4. B. Capel, Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* **18**, 675–689 (2017). doi: 10.1038/nrg.2017.60; pmid: 28804140
5. T. M. Williams, S. B. Carroll, Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nat. Rev. Genet.* **10**, 797–804 (2009). doi: 10.1038/nrg2687; pmid: 19834484
6. C. Weber, B. Capel, Sex reversal. *Curr. Biol.* **28**, R1234–R1236 (2018). doi: 10.1016/j.cub.2018.09.043; pmid: 30399341
7. D. Zhao *et al*., Somatic sex identity is cell autonomous in the chicken. *Nature* **464**, 237–242 (2010). doi: 10.1038/nature08852; pmid: 20220842
8. S. Naqvi *et al*., Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, eaaw7317 (2019). doi: 10.1126/science.aaw7317; pmid: 31320509
9. M. Oliva *et al*., The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020). doi: 10.1126/science.aba3066; pmid: 32913072
10. X. Yang *et al*., Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006). doi: 10.1101/gr.5217506; pmid: 16825664
11. J. E. Mank, The transcriptional architecture of phenotypic dimorphism. *Nat. Ecol. Evol.* **1**, 6 (2017). doi: 10.1038/s41559-016-0006; pmid: 28812569
12. P. J. Hansen, K. B. Dobbs, A. C. Denicol, L. G. B. Siqueira, Sex and the preimplantation embryo: Implications of sexual dimorphism in the preimplantation period for maternal programming of embryonic development. *Cell Tissue Res.* **363**, 237–247 (2016). doi: 10.1007/s00441-015-2287-4; pmid: 26391275
13. N. Engel, Sex differences in early embryogenesis: Inter-chromosomal regulation sets the stage for sex-biased gene networks. *BioEssays* **40**, e1800073 (2018). doi: 10.1002/bies.201800073; pmid: 29943439
14. P. W. Harrison *et al*., Sexual selection drives evolution and rapid turnover of male gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4393–4398 (2015). doi: 10.1073/pnas.1501339112; pmid: 25831521
15. A. Catalán, A. Macias-Muñoz, A. D. Briscoe, Evolution of Sex-Biased Gene Expression and Dosage Compensation in the Eye and Brain of Heliconius Butterflies. *Mol. Biol. Evol.* **35**, 2120–2134 (2018). doi: 10.1093/molbev/msy111; pmid: 29931127
16. G. Bontonou *et al*., "Evolution of chemosensory tissues and cells across ecologically diverse *Drosophilids*" *bioRxiv* (2023). doi:.doi: 10.1101/2023.04.14.536691
17. B. Reinius *et al*., An evolutionarily conserved sexual signature in the primate brain. *PLOS Genet.* **4**, e1000100 (2008). doi: 10.1371/journal.pgen.1000100; pmid: 18566661
18. H. Si *et al*., Human and murine kidneys show gender- and species-specific gene expression differences in response to injury. *PLOS ONE* **4**, e4802 (2009). doi: 10.1371/journal.pone.0004802; pmid: 19277126
19. D. F. Deegan, R. Karbalaei, J. Madzo, R. J. Kulathinal, N. Engel, The developmental origins of sex-biased expression in cardiac development. *Biol. Sex Differ.* **10**, 46 (2019). doi: 10.1186/s13293-019-0259-1; pmid: 31488212
20. M. Cardoso-Moreira *et al*., Gene expression across mammalian organ development. *Nature* **571**, 505–509 (2019). doi: 10.1038/s41586-019-1338-5; pmid: 31243369
21. S. Pantalacci, M. Sémon, Transcriptomics of developing embryos and organs: A raising tool for evo-devo. *J. Exp. Zool. B Mol. Dev. Evol.* **324**, 363–371 (2015). doi: 10.1002/jez.b.22595; pmid: 25387424

22. Materials and methods are available as supplementary materials.
23. J. A. M. Graves, Evolution of vertebrate sex chromosomes and dosage compensation. *Nat. Rev. Genet.* **17**, 33–46 (2016). doi: 10.1038/nrg.2015.2; pmid: 26616198
24. J. Hensman, N. D. Lawrence, M. Rattray, Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. *BMC Bioinformatics* **14**, 252 (2013). doi: 10.1186/1471-2105-14-252; pmid: 23962281
25. K. Munger, C. Baylis, Sex differences in renal hemodynamics in rats. *Am. J. Physiol.* **254**, F223–F231 (1988). doi: 10.1152/ajprenal.1988.254.2.F223; pmid: 3344806
26. A. Remuzzi, S. Puntorieri, A. Mazzoleni, G. Remuzzi, Sex related differences in glomerular ultrafiltration and proteinuria in Munich-Wistar rats. *Kidney Int.* **34**, 481–486 (1988). doi: 10.1038/ki.1988.206; pmid: 3199667
27. D. J. Waxman, M. G. Holloway, Sex differences in the expression of hepatic drug metabolizing enzymes. *Mol. Pharmacol.* **76**, 215–228 (2009). doi: 10.1124/mol.109.056705; pmid: 19483103
28. T. Groza *et al.*, The International Mouse Phenotyping Consortium: Comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038–D1045 (2023). doi: 10.1093/nar/gkac972; pmid: 36305825
29. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). doi: 10.1186/s13059-014-0550-8; pmid: 25516281
30. Y. Shao *et al.*, GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **29**, 682–696 (2019). doi: 10.1101/gr.238733.118; pmid: 30862647
31. M. Cardoso-Moreira *et al.*, Developmental gene expression differences between humans and mammalian models. *Cell Rep.* **33**, 108308 (2020). doi: 10.1016/j.celrep.2020.108308; pmid: 33113372
32. B. R. Johnson, Taxonomically restricted genes are fundamental to biology and evolution. *Front. Genet.* **9**, 407 (2018). doi: 10.3389/fgene.2018.00407; pmid: 30294344
33. M. J. Sheehan, P. Campbell, C. H. Miller, Evolutionary patterns of major urinary protein scent signals in house mice and relatives. *Mol. Ecol.* **28**, 3587–3601 (2019). doi: 10.1111/mec.15155; pmid: 31232499
34. H. Sezutsu, G. Le Goff, R. Feyereisen, Origins of P450 diversity. *Philos. Trans. R. Soc. London B Biol. Sci.* **368**, 20120428 (2013). doi: 10.1098/rstb.2012.0428; pmid: 23297351
35. D. C. Engelhart *et al.*, Systems biology analysis reveals eight SLC22 transporter subgroups, including OATs, OCTs, and OCTNs. *Int. J. Mol. Sci.* **21**, 1791 (2020). doi: 10.3390/ijms21051791; pmid: 32150922
36. D. W. Logan, T. F. Marton, L. Stowers, Species specificity in major urinary proteins by parallel evolution. *PLOS ONE* **3**, e3280 (2008). doi: 10.1371/journal.pone.0003280; pmid: 18815613
37. G. Gómez-Baena *et al.*, Molecular complexity of the major urinary protein system of the Norway rat, Rattus norvegicus. *Sci. Rep.* **9**, 10757 (2019). doi: 10.1038/s41598-019-46950-x; pmid: 31341188
38. A. Ransick *et al.*, Single-cell profiling reveals sex, lineage, and regional diversity in the mouse kidney. *Dev. Cell* **51**, 399–413.e7 (2019). doi: 10.1016/j.devcel.2019.10.005; pmid: 31689386
39. C. N. Goldfarb, K. Karri, M. Pyatkov, D. J. Waxman, Interplay between GH-regulated, sex-biased liver transcriptome and hepatic zonation revealed by single-nucleus RNA sequencing. *Endocrinology* **163**, bqac059 (2022). doi: 10.1210/endocr/bqac059; pmid: 35512247
40. A. N. Harris *et al.*, Differences in renal ammonia metabolism in male and female kidney. *Am. J. Physiol. Renal Physiol.* **315**, F211–F222 (2018). doi: 10.1152/ajprenal.00084.2018; pmid: 29561185
41. O. Oudar *et al.*, Differences in rat kidney morphology between males, females and testosterone-treated females. *Ren. Physiol. Biochem.* **14**, 92–102 (1991). doi: 10.1159/000173392; pmid: 1707550
42. R. Marcos *et al.*, Stereological assessment of sexual dimorphism in the rat liver reveals differences in hepatocytes and Kupffer cells but not hepatic stellate cells. *J. Anat.* **228**, 996–1005 (2016). doi: 10.1111/joa.12448; pmid: 26892301
43. A. N. Combes *et al.*, Single cell analysis of the developing mouse kidney provides deeper insight into marker gene expression and ligand-receptor crosstalk. *Development* **146**, dev178673 (2019). doi: 10.1242/dev.178673; pmid: 31118232
44. X. Wang *et al.*, Comparative analysis of cell lineage differentiation during hepatogenesis in humans and mice at the single-cell transcriptome level. *Cell Res.* **30**, 1109–1126 (2020). doi: 10.1038/s41422-020-0378-6; pmid: 32690901
45. S. Ma *et al.*, Caloric restriction reprograms the single-cell transcriptional landscape of *Rattus norvegicus* aging. *Cell* **180**, 984–1001.e22 (2020). doi: 10.1016/j.cell.2020.02.008; pmid: 32109414
46. C. M. McEvoy *et al.*, Single-cell profiling of healthy human kidney reveals features of sex-based transcriptional programs and tissue-specific immunity. *Nat. Commun.* **13**, 7634 (2022). doi: 10.1038/s41467-022-35297-z; pmid: 36496458
47. A. van Nas *et al.*, Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* **150**, 1235–1249 (2009). doi: 10.1210/en.2008-0563; pmid: 18974276
48. T. L. Conforto, Y. Zhang, J. Sherman, D. J. Waxman, Impact of CUX2 on the female mouse liver transcriptome: Activation of female-biased genes and repression of male-biased genes. *Mol. Cell. Biol.* **32**, 4611–4627 (2012). doi: 10.1128/MCB.00886-12; pmid: 22966202
49. D. J. Waxman, P. A. Ram, S.-H. Park, H. K. Choi, Intermittent plasma growth hormone triggers tyrosine phosphorylation and nuclear translocation of a liver-expressed, Stat 5-related DNA binding protein. Proposed role as an intracellular regulator of male-specific liver gene transcription. *J. Biol. Chem.* **270**, 13262–13270 (1995). doi: 10.1074/jbc.270.22.13262; pmid: 7768925
50. H. K. Choi, D. J. Waxman, Plasma growth hormone pulse activation of hepatic JAK-STAT5 signaling: Developmental regulation and role in male-specific liver gene expression. *Endocrinology* **141**, 3245–3255 (2000). doi: 10.1210/endo.141.9.7638; pmid: 10965895
51. J.-O. Jansson, S. Edén, O. Isaksson, Sexual dimorphism in the control of growth hormone secretion. *Endocr. Rev.* **6**, 128–150 (1985). doi: 10.1210/edrv-6-2-128; pmid: 2861084
52. J. N. MacLeod, N. A. Pampori, B. H. Shapiro, Sex differences in the ultradian pattern of plasma growth hormone concentrations in mice. *J. Endocrinol.* **131**, 395–399 (1991). doi: 10.1677/joe.0.1310395; pmid: 1783886
53. H. K. Choi, D. J. Waxman, Growth hormone, but not prolactin, maintains, low-level activation of STAT5a and STAT5b in female rat liver. *Endocrinology* **140**, 5126–5135 (1999). doi: 10.1210/endo.140.11.7106; pmid: 10537141
54. K. H. Clodfelter *et al.*, Sex-dependent liver gene expression is extensive and largely dependent upon signal transducer and activator of transcription 5b (STAT5b): STAT5b-dependent activation of male genes and repression of female genes revealed by microarray analysis. *Mol. Endocrinol.* **20**, 1333–1351 (2006). doi: 10.1210/me.2005-0489; pmid: 16469768
55. Y. Zhang, E. V. Laz, D. J. Waxman, Dynamic, sex-differential STAT5 and BCL6 binding to sex-biased, growth hormone-regulated genes in adult mouse liver. *Mol. Cell. Biol.* **32**, 880–896 (2012). doi: 10.1128/MCB.06312-11; pmid: 22158971
56. R. D. Meyer, E. V. Laz, T. Su, D. J. Waxman, Male-specific hepatic Bcl6: Growth hormone-induced block of transcription elongation in females and binding to target genes inversely coordinated with STAT5. *Mol. Endocrinol.* **23**, 1914–1926 (2009). doi: 10.1210/me.2009-0242; pmid: 19797429
57. E. V. Laz, M. G. Holloway, C.-S. Chen, D. J. Waxman, Characterization of three growth hormone-responsive transcription factors preferentially expressed in adult female liver. *Endocrinology* **148**, 3327–3337 (2007). doi: 10.1210/en.2006-1192; pmid: 17412818
58. T. L. Conforto, G. F. Steinhardt4th, D. J. Waxman, Cross talk between GH-regulated transcription factors HNF6 and CUX2 in adult mouse liver. *Mol. Endocrinol.* **29**, 1286–1302 (2015). doi: 10.1210/me.2015-1028; pmid: 26218442
59. S. Gao *et al.*, Androgen receptor tumor suppressor function is mediated by recruitment of retinoblastoma protein. *Cell Rep.* **17**, 966–976 (2016). doi: 10.1016/j.celrep.2016.09.064; pmid: 27760327
60. G. Gritsina, W.-Q. Gao, J. Yu, Transcriptional repression by androgen receptor: Roles in castration-resistant prostate cancer. *Asian J. Androl.* **21**, 215–223 (2019). doi: 10.4103/aja.aja_19_19; pmid: 30950412
61. J. S. Carroll *et al.*, Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289–1297 (2006). doi: 10.1038/ng1901; pmid: 17013392
62. H. U. Osmanbeyoglu *et al.*, Estrogen represses gene expression through reconfiguring chromatin structures. *Nucleic Acids Res.* **41**, 8061–8071 (2013). doi: 10.1093/nar/gkt586; pmid: 23821662
63. M. Blencowe *et al.*, Relative contributions of sex hormones, sex chromosomes, and gonads to sex differences in tissue gene regulation. *Genome Res.* **32**, 807–824 (2022). doi: 10.1101/gr.275965.121; pmid: 35396276
64. G. Ling, A. Sugathan, T. Mazor, E. Fraenkel, D. J. Waxman, Unbiased, genome-wide in vivo mapping of transcriptional regulatory elements reveals sex differences in chromatin structure associated with sex-specific liver gene expression. *Mol. Cell. Biol.* **30**, 5531–5544 (2010). doi: 10.1128/MCB.00601-10; pmid: 20876297
65. A. Sugathan, D. J. Waxman, Genome-wide analysis of chromatin states reveals distinct mechanisms of sex-dependent gene regulation in male and female mouse liver. *Mol. Cell. Biol.* **33**, 3594–3610 (2013). doi: 10.1128/MCB.00280-13; pmid: 23836885
66. K. Hyun, J. Jeon, K. Park, J. Kim, Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324–e324 (2017). doi: 10.1038/emm.2017.11; pmid: 28450737
67. W.-J. Ma, P. Veltsos, R. Sermier, D. J. Parker, N. Perrin, Evolutionary and developmental dynamics of sex-biased gene expression in common frogs with proto-Y chromosomes. *Genome Biol.* **19**, 156 (2018). doi: 10.1186/s13059-018-1548-4; pmid: 30290841
68. J. Djordjevic, Z. Dumas, M. Robinson-Rechavi, T. Schwander, D. J. Parker, Dynamics of sex-biased gene expression during development in the stick insect Timema californicum. *Heredity* **129**, 113–122 (2022). doi: 10.1038/s41437-022-00536-y; pmid: 35581477
69. L. Shi, Z. Zhang, B. Su, Sex biased gene expression profiling of human brains at major developmental stages. *Sci. Rep.* **6**, 21181 (2016). doi: 10.1038/srep21181; pmid: 26880485
70. G. Shaw, M. B. Renfree, R. V. Short, W.-S. O, Experimental manipulation of sexual differentiation in wallaby pouch young treated with exogenous steroids. *Development* **104**, 689–701 (1988). doi: 10.1242/dev.104.4.689; pmid: 3268410
71. M. B. Renfree, R. V. Short, Sex determination in marsupials: Evidence for a marsupial-eutherian dichotomy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **322**, 41–53 (1988). pmid: 2907802
72. J. E. Deakin, Implications of monotreme and marsupial chromosome evolution on sex determination and differentiation. *Gen. Comp. Endocrinol.* **244**, 130–138 (2017). doi: 10.1016/j.ygcen.2015.09.029; pmid: 26431612
73. M. B. Renfree, K. Y. Chew, G. Shaw, Hormone-independent pathways of sexual differentiation. *Sex Dev.* **8**, 327–336 (2014). pmid: 24577198
74. X. Zhang *et al.*, A novel nonsense mutation of PHF6 in a female with extended phenotypes of Borjeson-Forssman-Lehmann syndrome. *J. Clin. Res. Pediatr. Endocrinol.* **11**, 419–425 (2019). doi: 10.4274/jcrpe.galenos.2019.2018.0220; pmid: 30630810
75. J. Carrillo *et al.*, High resolution melting analysis for the identification of novel mutations in DKC1 and TERT genes in patients with dyskeratosis congenita. *Blood Cells Mol. Dis.* **49**, 140–146 (2012). doi: 10.1016/j.bcmd.2012.05.008; pmid: 22664374
76. T. Connallon, L. L. Knowles, Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet.* **21**, 495–499 (2005). doi: 10.1016/j.tig.2005.07.006; pmid: 16039005
77. P. Innocenti, E. H. Morrow, The sexually antagonistic genes of *Drosophila melanogaster*. *PLOS Biol.* **8**, e1000335 (2010). doi: 10.1371/journal.pbio.1000335; pmid: 20305719
78. E. Y. Shen *et al.*, Epigenetics and sex differences in the brain: A genome-wide comparison of histone-3 lysine-4 trimethylation (H3K4me3) in male and female mice. *Exp. Neurol.* **268**, 21–29 (2015). doi: 10.1016/j.expneurol.2014.08.006; pmid: 25131640
79. L. E. Cabrera Zapata *et al.*, X-linked histone H3K27 demethylase Kdm6a regulates sexually dimorphic differentiation of hypothalamic neurons. *Cell. Mol. Life Sci.* **78**, 7043–7060 (2021). doi: 10.1007/s00018-021-03945-0; pmid: 34633482
80. J. C. Link *et al.*, X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. *J. Clin. Invest.* **130**, 5688–5702 (2020). doi: 10.1172/JCI140223; pmid: 32701509
81. W. Ma *et al.*, Sex-biased and parental allele-specific gene regulation by KDM6A. *Biol. Sex Differ.* **13**, 40 (2022). doi: 10.1186/s13293-022-00452-0; pmid: 35871105
82. A. Michna, "splineTimeR: Time-course differential gene expression data analysis using spline regression models followed by gene association network reconstruction"

(R package version 1.28.0, 2023); https://doi.org/10.18129/B9.bioc.splineTimeR.
83. A. Conesa, M. J. Nueda, "maSigPro: Significant gene expression profile differences in time course gene expression data" (R package version 1.72.0, 2023); https://doi.org/10.18129/B9.bioc.maSigPro.
84. D. Gerard, Data-based RNA-seq simulations by binomial thinning. *BMC Bioinformatics* **21**, 206 (2020). doi: 10.1186/s12859-020-3450-9; pmid: 32448189
85. G. X. Y. Zheng *et al*., Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: 10.1038/ncomms14049; pmid: 28091601
86. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi: 10.1038/nbt.4096; pmid: 29608179
87. G. Yu, L.-G. Wang, Q.-Y. He, ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015). doi: 10.1093/bioinformatics/btv145; pmid: 25765347
88. P. Pihlajamaa *et al*., Tissue-specific pioneer factors associate with androgen receptor cistromes and transcription programs. *EMBO J.* **33**, 312–326 (2014). doi: 10.1002/embj.201385895; pmid: 24451200
89. J. Gertz *et al*., Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol. Cell* **52**, 25–36 (2013). doi: 10.1016/j.molcel.2013.08.037; pmid: 24076218
90. Z. Li, G. Tuteja, J. Schug, K. H. Kaestner, Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer. *Cell* **148**, 72–83 (2012). doi: 10.1016/j.cell.2011.11.026; pmid: 22265403
91. Code for: L. Rodríguez-Montes, S. Ovchinnikova, X. Yuan, T. Studer, I. Sarropoulos, S. Anders, H. Kaessmann, M. Cardoso-Moreira, Sex-biased gene expression across mammalian organ development and evolution, Zenodo (2023); https://doi.org/10.5281/zenodo.8199053.

ACKNOWLEDGMENTS

We thank R. Arguello, members of the Kaessmann group, and members of the Cardoso-Moreira group for discussion and comments on the manuscript. **Funding:** This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (grant CC2185 to M.C.-M.), the UK Medical Research Council (grant CC2185 to M.C.-M.), and the Wellcome Trust (grant CC2185 to M.C.-M.); the European Research Council (grant 615253 to OntoTransEvol to H.K.); the Klaus Tschira Foundation (S.A.); and Deutsche Forschungsgemeinschaft (grant DFG SFB 1366 to S.O. and DFG grant KA 1710/4-1 to H.K.).
Author contributions: M.C.-M. and H.K. conceived the study. L.R.-M. performed most analyses with contributions from X.Y., T.S., I.S., S.O., and S.A. S.O. and S.A. developed the in-house pipeline to detect sex-biased genes. X.Y. generated and analyzed snRNA-seq data. L.R.-M. and M.C.-M. wrote the manuscript with contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The raw and processed data generated in this study were deposited in ArrayExpress with the accession code E-MTAB-12180. All other data are provided in the manuscript or in the supplementary materials (*22*). The code used to analyze the data is available at https://github.com/Leticia314/Sex_bias_manuscript and is also archived at Zenodo (*91*). Processed data can be interactively explored at https://apps.kaessmannlab.org/sexbiasapp. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by the Wellcome Trust (grant CC2185), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf1046
Materials and Methods
Figs. S1 to S9
Tables S1 to S16
References (*92*–*122*)
MDAR Reproducibility Checklist

Submitted 29 September 2022; resubmitted 31 May 2023
Accepted 18 September 2023
10.1126/science.adf1046

RESEARCH ARTICLE

IMMUNOLOGY

CD8 T cell tolerance results from eviction of immature autoreactive cells from the thymus

Mohamed Elsherif Badr*, Zhongmei Zhang, Xuguang Tai, Alfred Singer*

CD8 T cell tolerance is thought to result from clonal deletion of autoreactive thymocytes before they differentiate into mature CD8 T cells in the thymus. However, we report that, in mice, CD8 T cell tolerance instead results from premature thymic eviction of immature autoreactive CD8 thymocytes into the periphery, where they differentiate into self-tolerant mature CD8 T cells. Premature thymic eviction is triggered by T cell receptor (TCR)–driven down-regulation of the transcriptional repressor Gfi1, which induces expression of sphingosine-1–phosphate receptor-1 (S1P1) on negatively selected immature CD8 thymocytes. Thus, premature thymic eviction is the basis for CD8 T cell tolerance and is the mechanism responsible for the appearance in the periphery of mature CD8 T cells bearing autoreactive TCRs that are absent from the thymus.

Selection of a functional and self-tolerant T cell receptor (TCR) repertoire in the thymus requires that thymocytes undergo positive and negative selection on the basis of the specificity of their TCRs for self-ligands. Thymocytes that engage self-ligands with low affinity undergo positive selection into mature CD4 or CD8 T cells, whereas thymocytes with potentially autoreactive TCRs engage self-ligands with high affinity and undergo negative selection (*1*). The major consequence of negative selection is thought to be clonal deletion, which eliminates autoreactive thymocytes before their differentiation into mature T cells (*2*–*4*). However, it is now understood that TCR signaling is sustained during major histocompatibility class II (MHC-II)–specific CD4 T cell development but is disrupted during major histocompatibility class I (MHC-I)–specific CD8 T cell development (*5*–*7*). Accordingly, MHC-I–specific TCR signaling may be disrupted before deleting enough autoreactive thymocytes to induce CD8 T cell tolerance. In this work, we report that CD8 T cell tolerance does not result from clonal deletion but is induced by the premature eviction of immature autoreactive thymocytes into the periphery, where they complete their differentiation into self-tolerant mature CD8 T cells.

MHC-I–specific autoreactive thymocytes survive negative selection

Autoreactive thymocytes that avoid clonal deletion during negative selection express PD-1, an attenuator of TCR signaling, and become TCR^{hi}PD-1^{+} cells (*8*–*10*). TCR^{hi}PD-1^{+} cells were significantly higher in both frequency and PD-1 expression among MHC-I–selected compared with MHC-II–selected thymocytes from *H-2Ab1*$^{-/-}$ (MHC-II^{-}) and *b2m*$^{-/-}$ (MHC-I^{-}) mice, respectively (Fig. 1, A and B). These TCR^{hi}PD-1^{+} thymocytes lacked cleaved caspase-3, which confirmed that they were not dying cells (Fig. 1, A and B). Thus, a significantly higher frequency of MHC-I–specific compared with MHC-II–specific autoreactive thymocytes survive negative selection.

To compare developing MHC-I– and MHC-II–specific thymocytes with TCRs that recognize the same self-antigen, we examined HY^{cd4} and Marilyn mice, whose T cells both express transgenic TCRs specific for the HY male antigen and induce negative selection in male mice but positive selection in female mice (*11*, *12*). The transgenic TCRs in HY^{cd4} mice are clonotype T3.70^{+} and engage MHC-I ligands to generate CD8 T cells (*11*), whereas the transgenic TCRs in Marilyn mice are TCR-Vβ6^{+} and engage MHC-II ligands to generate CD4 T cells (*12*). Notably, in male mice, Marilyn thymocytes failed to become TCR^{hi}, whereas HY^{cd4} thymocytes became TCR^{hi} (Fig. 1, C and D). These TCR^{hi} thymocytes were generated by the transgenic TCR because their thymocyte profiles were identical in *Rag2*-sufficient and *Rag2*-deficient HY^{cd4} mice (Fig. 1, C and D). Thus, negative selection does not prevent MHC-I–selected thymocytes from differentiating into TCR^{hi} cells.

MHC-I–selected thymocytes that were TCR^{hi} in HY^{cd4} male (HY^{cd4}M) mice did not become CD8–single-positive (CD8SP) but remained CD4/8^{dull} (Fig. 1, E and F). The absence of CD8SP thymocytes in negatively selecting HY^{cd4}M mice has been attributed to clonal deletion (*11*, *13*). However, this conclusion was contradicted by the presence of HY-specific CD8αβ T cells in the periphery of male mice in greater numbers compared with HY^{cd4} female (HY^{cd4}F) mice (Fig. 1, E and F, and fig. S1A). By contrast, HY-specific CD4 T cells were not present in the periphery of Marilyn male mice (fig. S1B). Thus, unlike MHC-II–selected autoreactive CD4 T cells, MHC-I–selected autoreactive CD8 T cells survive negative selection and appear in the periphery.

To assess cell death during MHC-I–specific negative selection, we examined cleaved caspase-3 in HY^{cd4}M thymocytes (Fig. 1G). Less than 1% contained cleaved caspase-3, and these dying cells were Bcl-2^{lo}CCR7^{lo}. An even higher frequency of negatively selected thymocytes were Bcl2^{+}CCR7^{hi} and cleaved–caspase-3^{-} and were differentiating, not dying, cells (Fig. 1G). Because thymocyte clonal deletion during negative selection is induced by B7 costimulatory ligands (CD80 and CD86) (*8*, *14*, *15*), we compared CD8 T cell numbers in HY^{cd4}M mice with and without B7 deficiency. Impaired clonal deletion in B7^{DKO}HY^{cd4}M mice increased TCR^{hi} thymocyte numbers by only one-third (from 14 million to 21 million) and did not reconstitute generation of CD8SP thymocytes (Fig. 1, H and I). Thus, clonal deletion eliminated only a minority of autoreactive MHC-I–specific thymocytes and failed to explain the absence of CD8SP cells in the thymus.

Immature autoreactive CD8 thymocytes are evicted from the thymus

To understand how HY-specific CD8 T cells appeared in the periphery when they were absent from the male thymus, we characterized the TCR^{hi}CD4/8^{dull} thymocytes that populated the male thymus (Fig. 2A). Female thymi contained mature CD8SP thymocytes, whereas TCR^{hi}CD4/8^{dull} thymocytes in male thymi were PD-1^{+}Bcl-2^{+} (indicating that they had been strongly TCR signaled) and were CD24^{hi}Qa2^{-} (indicating that they were immature) (Fig. 2A). Nevertheless, immature TCR^{hi}CD4/8^{dull} male thymocytes were CD8 lineage–committed because they expressed the cytotoxic-lineage transcription factor Runx3d, albeit at lower levels compared with CD8SP thymocytes (Fig. 2B).

To determine whether immature TCR^{hi}CD4/8^{dull} thymocytes exited the male thymus, we examined thymocytes 5 min after intravenous (i.v.) injection of anti-CD5 monoclonal antibody (mAb), which only had enough time to bind to exiting thymocytes that contacted the blood circulation (*16*). Whereas CD5(i.v.)$^{+}$ cells exiting B6 thymi were single positive (SP), CD5(i.v.)$^{+}$ thymocytes exiting HY^{cd4}M thymi were T3.70^{hi}CD4/8^{dull} (Fig. 2C). Such T3.70^{hi}CD4/8^{dull} cells were also present in the periphery of male mice, and these peripheral T3.70^{hi}CD4/8^{dull} cells resembled immature thymocytes in being CD24^{hi}PD-1^{+} (Fig. 2, D and E). Thus, TCR^{hi}CD4/8^{dull} thymocytes exit the male thymus despite being immature.

Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA.
*Corresponding author. Email: sherif.badr@nih.gov (M.E.B.); singera@nih.gov (A.S.)

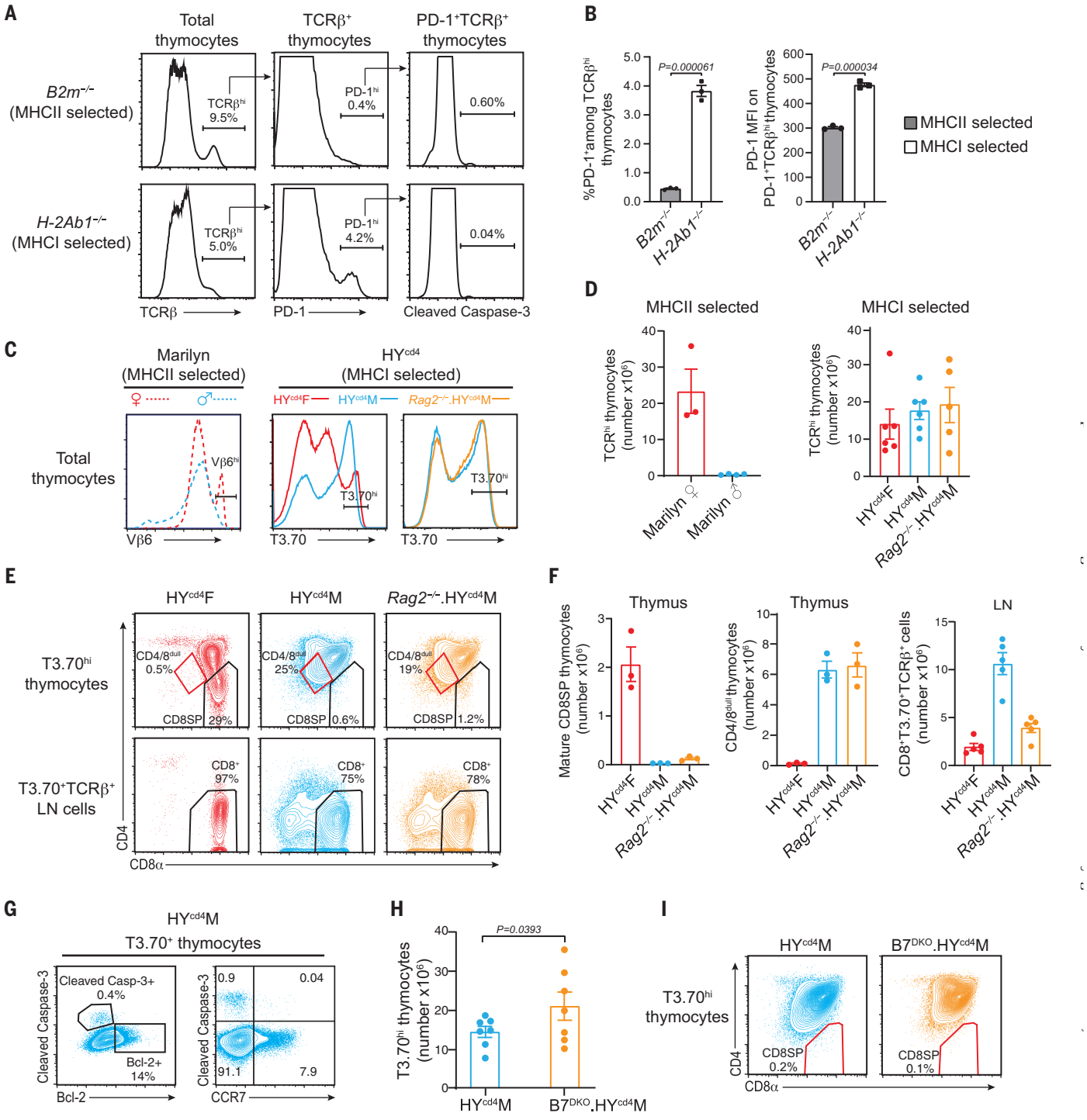


Fig. 1. Characterization of developing thymocytes and T cells that preferentially survive strong TCR signaling during negative selection in the thymus. (**A**) Thymocyte profiles from mice containing either MHC-II–selected or MHC-I–selected T cells. Shown are TCR profiles of total thymocytes (left), PD-1 profiles on TCR^{hi} (CD1d–PBS57^{-} non-iNKT) thymocytes (middle), and cleaved caspase-3 expression in PD-1^{+}TCR^{hi} (non-iNKT) thymocytes (right). (**B**) Percent PD-1^{+} cells among TCRβ^{hi} thymocytes (left) and quantification of their surface PD-1 expression as mean fluorescence intensity (MFI) (right). (**C**) TCR profiles of male versus female thymocytes from Marilyn and HY^{cd4} transgenic mice. (**D**) TCR^{hi} thymocyte numbers in Marilyn and HY^{cd4} transgenic mice. (**E**) CD4 versus CD8α expression on TCR^{hi} thymocytes (top) and TCR^{+} lymph node (LN) cells (bottom) in male and female HY^{cd4} mice. (**F**) Thymocyte and LN T cell numbers in HY^{cd4} transgenic mice. (**G**) Expression of cleaved caspase-3 versus Bcl-2 (left) or CCR7 (right) in HY-specific thymocytes from HY^{cd4} male mice. (**H**) TCR^{hi} thymocyte numbers in WT and B7-deficient [B7 double-knockout (B7^{DKO})] HY^{cd4} male mice. (**I**) CD4 versus CD8α expression on TCR^{hi} thymocytes on WT and B7-deficient HY^{cd4} male mice. [(A) to (C), (E), (G), and (I)] Representative data are from three independent experiments. [(D), (F), and (H)] Data are from two to four independent experiments. *P* values were determined using unpaired two-tailed Student's *t* tests (means ± SEMs).

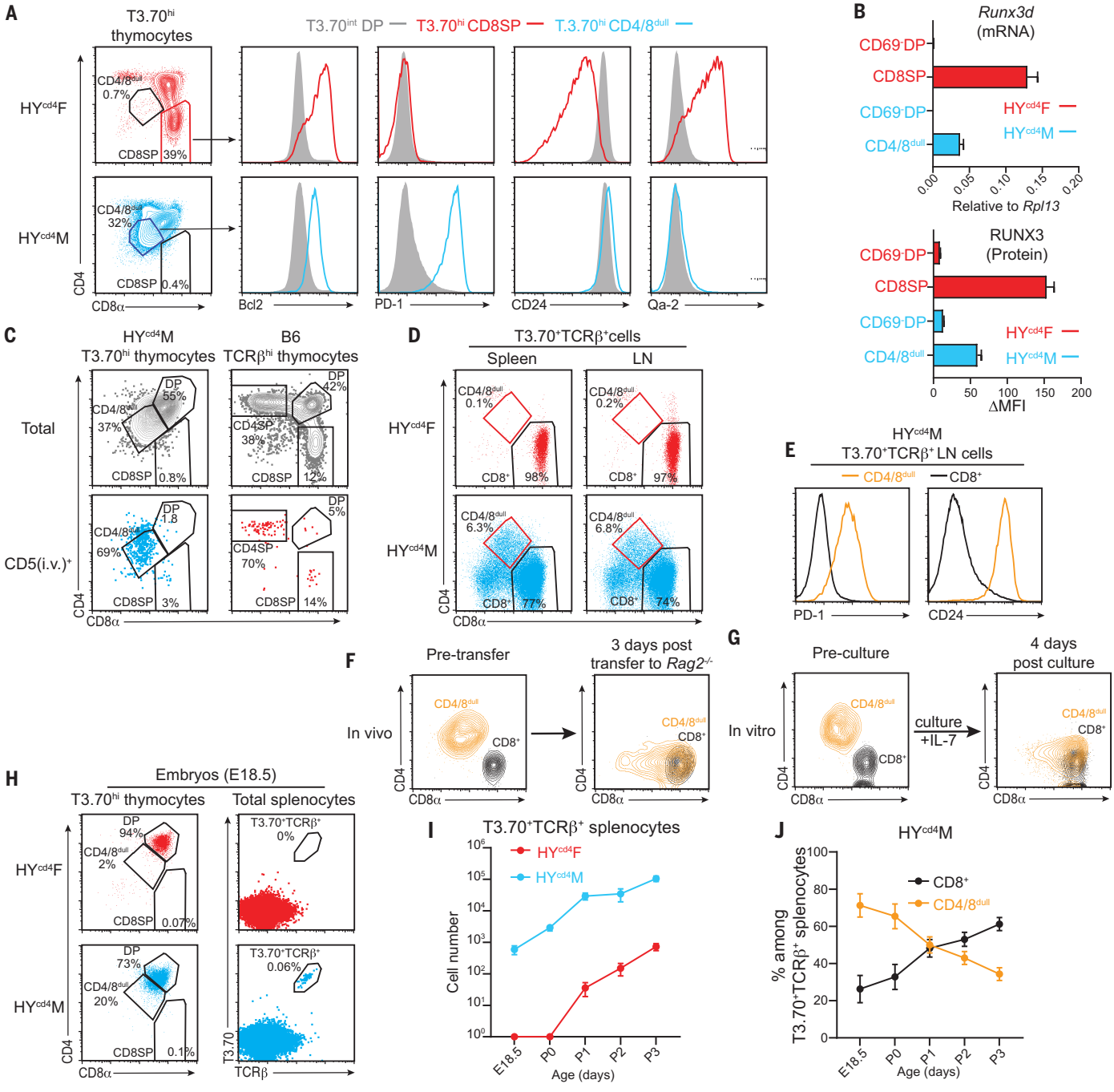


Fig. 2. Clonal eviction of autoreactive CD8 thymocytes signaled to undergo negative selection. (**A**) CD4 versus CD8α expression on TCR^{hi} thymocytes from HY^{cd4} male and female mice (left) and expression of various proteins in the indicated thymocyte subsets (single-color histograms). (**B**) Expression of *Runx3d* mRNA (top) and RUNX3 protein (bottom) in thymocyte subsets from HY^{cd4} male and female mice. (**C**) To identify cells exiting the thymus, thymocytes were harvested 5 min after in vivo i.v. injection of phycoerythrin (PE)–conjugated anti-CD5 mAb. Thymocytes that bound the injected anti-CD5 mAb were referred to as CD5(i.v.)$^+$ cells. (**D**) CD4 versus CD8 profiles of HY-specific T3.70$^+$ peripheral T cells in HY^{cd4} male and female mice. Only negatively selecting male mice contained peripheral CD4/8^{dull} HY-specific T cells. (**E**) PD-1 and CD24 expression on peripheral CD4/8^{dull} and CD4$^-$8$^+$ HY-specific T cells in HY^{cd4} male mice. (**F**) CD4/8^{dull} and CD8$^+$ HY-specific T3.70$^+$ T cells were sorted from the spleen and LN of HY^{cd4} male mice and were adoptively transferred into separate *Rag2*$^{-/-}$ hosts. CD4 versus CD8α expression on each transferred cell population was overlayed before and 3 days after transfer into host mice. (**G**) CD4/8^{dull} and CD8$^+$ HY-specific T3.70$^+$ T cells were sorted from the spleen and LN of HY^{cd4} male mice and were placed into separate in vitro cultures containing IL-7 for 4 days. Overlayed CD4 versus CD8α expression on sorted T cells before and after IL-7 culture is shown. (**H**) CD4 versus CD8α expression on HY-specific T3.70^{hi} thymocytes and splenocytes in male and female embryonic day 18.5 (E18.5) HY^{cd4} mice. (**I**) Comparison of HY-specific T3.70$^+$ T cells in HY^{cd4} male versus female neonatal mice at indicated ages. P0, postnatal day 0. (**J**) Frequencies of CD4/8^{dull} and CD8$^+$ cells among T3.70$^+$TCRβ$^+$ splenocytes in HY^{cd4}M neonatal mice. [(A) and (C) to (G)] Data are representative of three or four independent experiments. (B) Data are from three independent experiments with technical triplicates. (H) Representative data are from three or four HY^{cd4} embryos of each gender. [(I) and (J)] n = 3 to 8 from each gender at each time point. Data are expressed as means ± SEMs.

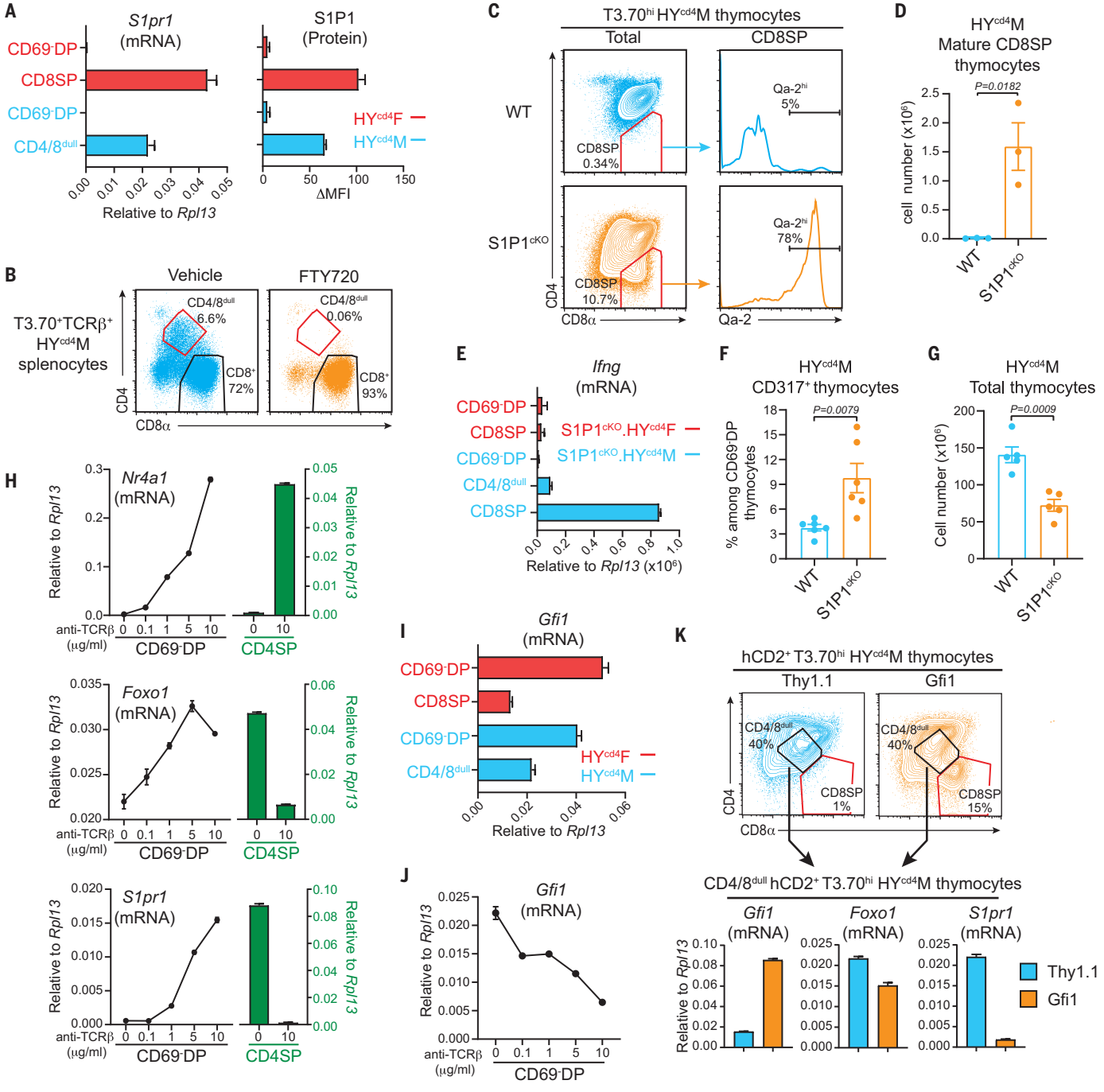


Fig. 3. Strong TCR signaling of immature thymocytes induces S1P1 expression and clonal eviction by down-regulating Gfi1. (**A**) *S1pr1* mRNA (left) and S1P1 protein (right) expression in HY^{cd4} male and female thymocytes. (**B**) CD4 versus CD8α profiles of HY-specific T3.70^{+} splenocytes in HY^{cd4} male mice after five daily injections of either FTY720 or vehicle. (**C**) CD4 versus CD8α and Qa-2 profiles of HY-specific T3.70^{+} TCR^{hi} thymocytes from S1P1-deficient [S1P1 conditional knockout (S1P1^{cKO})] and WT HY^{cd4} male mice. (**D**) Numbers of mature (CD24^{lo}) CD8SP thymocytes in S1P1^{cKO} and WT HY^{cd4} male mice. (**E**) *Ifng* mRNA in thymocyte subsets from S1P1^{cKO} and WT HY^{cd4} male and female mice. (**F**) Frequency of CD317^{+} preselection (CD69^{-}DP) thymocytes in S1P1^{cKO} and WT HY^{cd4} male mice. (**G**) Number of total thymocytes in S1P1^{cKO} and WT HY^{cd4} male mice. (**H**) *Nr4a1*, *Foxo1*, and *S1pr1* mRNA expression in immature (CD69^{-}DP) and mature (CD4SP) B6 thymocytes after overnight stimulation with immobilized plate-bound anti-TCRβ. (**I**) *Gfi1* mRNA expression in HY^{cd4} male and female thymocytes. (**J**) *Gfi1* mRNA expression in immature (CD69^{-}DP) B6 thymocytes after overnight stimulation with immobilized plate-bound anti-TCRβ. (**K**) Donor HY^{cd4}M bone marrow cells (containing CD4-cre) were infected in vitro with FLEX/Gfi1 or FLEX/Thy1.1 lentiviruses, and the lentivirus-infected bone marrow cells were then injected into lethally irradiated B6 CD45.1 host mice. Donor-origin HY^{cd4}M thymocytes were analyzed 4 to 6 weeks later, and those constitutively expressing Gfi1 or Thy1.1 lentiviral proteins were identified by surface expression of hCD2 reporter protein. [(A), (E), and (H) to (K)] Representative experiment from two to four independent experiments with technical triplicates. [(B) to (D)] Representative experiment from three or four independent experiments. [(F) and (G)] Data are from two to three independent experiments. *P* values were determined using unpaired two-tailed Student's *t* tests. Data are expressed as means ± SEMs.

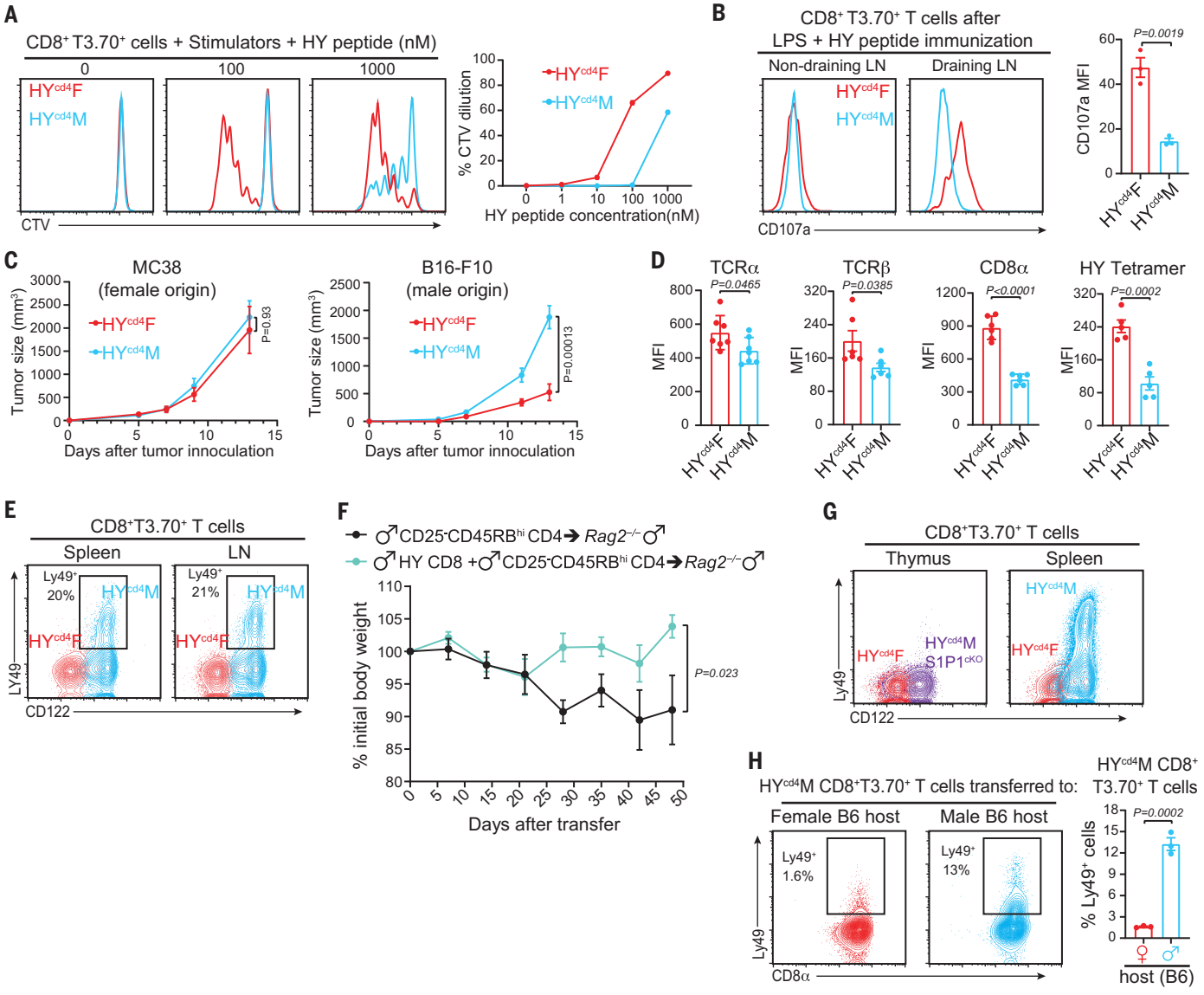


Fig. 4. Peripherally generated CD8 T cells are self-tolerant. (**A**) In vitro proliferative responses of CD8 T cells from HY^{cd4} male and female mice to 3 days of stimulation with varying doses of HY peptide presented by B6 splenocytes, as measured by the dilution of cell trace violet (CTV). (**B**) In vivo response as measured by CD107a expression of CD8 T cells from draining and nondraining popliteal LNs in HY^{cd4} male and female mice to footpad injection with HY peptide and LPS 16 hours earlier. (**C**) In vivo growth of female (MC38) or male (B16-F10) tumor cells subcutaneously injected into HY^{cd4} mice. (**D**) Surface TCR and CD8 expression and quantification of HY tetramer binding to CD8 LN T cells from HY^{cd4} male and female mice. (**E**) Ly49 versus CD122 profiles of peripheral CD8 T cells from HY^{cd4} male and female mice. (**F**) Relative body weights of male Rag2$^{-/-}$ mice after transfer of naïve CD4^{+}CD25^{-}CD45RB^{hi} B6 male T cells alone or together with male HY^{cd4} CD8 T cells. (**G**) Ly49 versus CD122 profiles of CD8SP thymocytes or CD8 spleen T cells from HY^{cd4}F mice, S1P1^{cKO}.HY^{cd4}M mice, and HY^{cd4}M mice. (**H**) Ly49 versus CD8α expression on CD8^{+}T3.70^{+} T cells from HY^{cd4}M mice analyzed 7 days after adoptive transfer into male and female B6 hosts. [(A), (B), and (E)] Representative data are from three independent experiments. (D) Data are pooled from three independent experiments. [(C) and (F)] Representative experiment from two independent experiments (n = 5 mice from each group). P values were determined using unpaired two-tailed Student's t tests except for (C) and (F), where two-way analysis of variance (ANOVA) was used instead (means ± SEMs).

To determine whether immature TCR^{hi}CD4/8^{dull} cells complete their differentiation in the periphery, we adoptively transferred them into immunodeficient Rag2$^{-/-}$ hosts and found that they became CD8 T cells (Fig. 2F). Immature CD4/8^{dull} cells also became CD8 T cells in interleukin-7 (IL-7) in vitro cultures, and their differentiation into CD8 T cells depended on Runx3d and cytokine receptor gamma chain (γc) (Fig. 2G and fig. S2), as does CD8 T cell generation in the normal thymus (*17*). Thus, during negative selection, MHC-I–selected thymocytes became immature TCR^{hi}CD4/8^{dull} cells that express PD-1 and Runx3d and that prematurely exit the thymus to complete their differentiation into mature CD8 T cells in the periphery. We refer to the exit of immature autoreactive thymocytes from the thymus into the periphery as clonal eviction.

Premature clonal eviction of autoreactive CD4/8^{dull} thymocytes also occurred during normal ontogeny, as T3.70^{+} cells appeared in the periphery of negatively selecting male mice 2 days earlier than in positively selecting

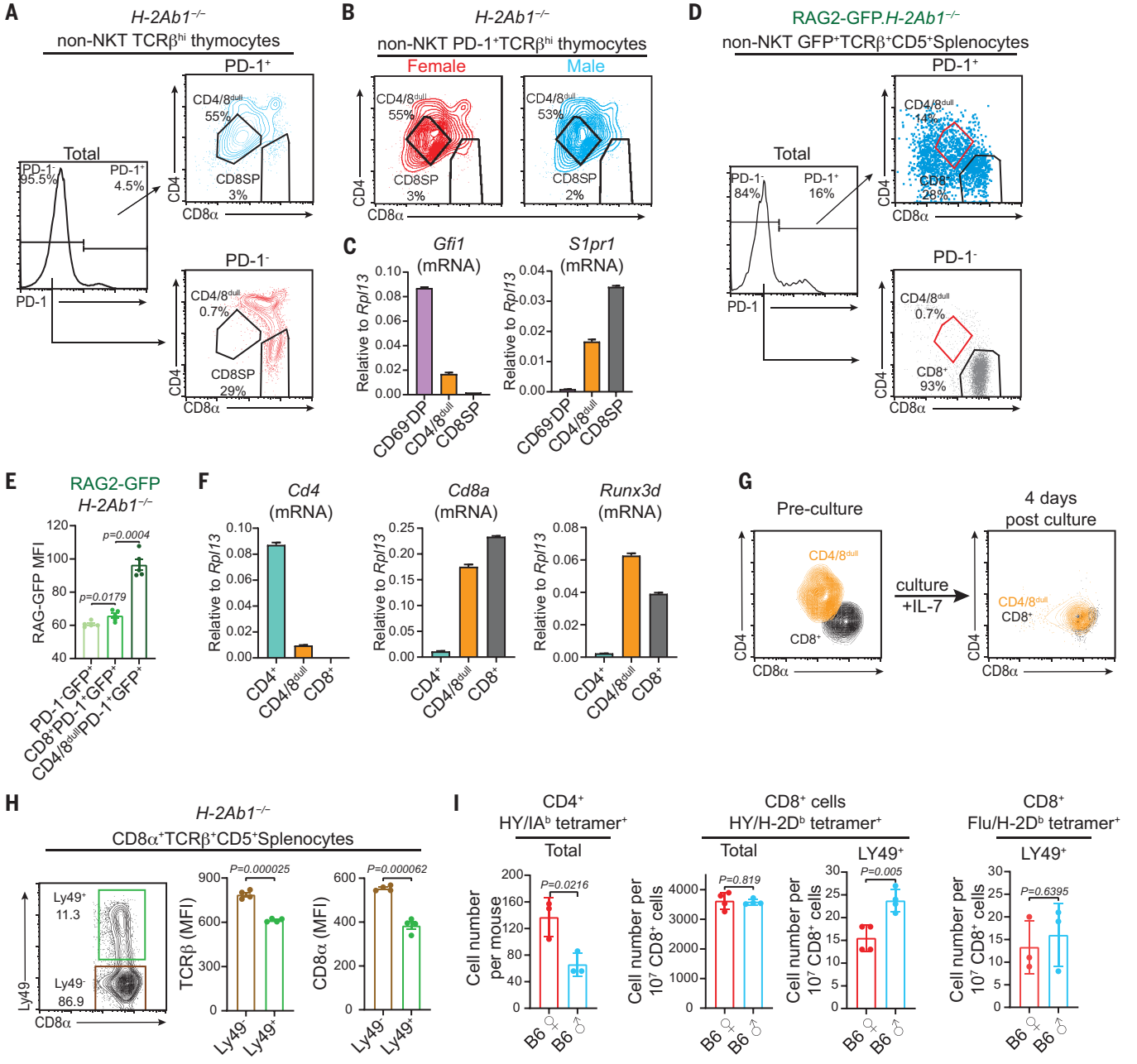


Fig. 5. Assessment of clonal eviction in normal polyclonal mice. (**A**) Analysis of MHC-I–selected thymocytes from MHC-II^{-} polyclonal mice. PD-1 expression on TCR^{hi} (non-NKT) thymocytes (left), and CD4 versus CD8α profiles on PD-1^{+} and PD-1^{-} TCR^{hi} thymocytes (right). (**B**) CD4 versus CD8α profiles on PD-1^{+}TCRβ^{hi} thymocytes from MHC-II^{-} male and female mice. (**C**) *Gfi1* and *S1pr1* mRNA expression in sorted thymocyte subsets from MHC-II^{-} mice. (**D**) Analysis of RTEs in the spleens of RAG2-GFP.MHC-II^{-} mice. PD-1 expression (left) and CD4 versus CD8α profiles on PD-1^{+} and PD-1^{-} RTEs (right). (**E**) Quantification of RAG-GFP fluorescence in spleen RTEs from RAG2-GFP.MHC-II^{-} mice. (**F**) Expression of *Cd4*, *Cd8a*, and *Runx3d* mRNA in sorted spleen and LN T cell subsets in Rag-GFP.MHC-II^{-} mice. (**G**) RTE CD4/8^{dull}PD-1^{+} and CD8^{+}PD-1^{+} were sorted from the spleens and LNs of RAG-GFP.MHC-II^{-} mice and placed into separate IL-7 in vitro cultures for 4 days. (**H**) Ly49 versus CD8α profile of total CD8 T cells from MHC-II^{-} spleens (left) and comparison of TCRβ and CD8α expression on Ly49^{-} and Ly49^{+} CD8 T cell subsets (bar graphs). (**I**) Comparison between B6 male and B6 female mice of HY–IA^{b} tetramer binding to CD4 T cells (left), HY–H-2D^{b} tetramer binding to total and Ly49^{+} CD8 T cells (middle), and Flu–H-2D^{b} tetramer binding to Ly49^{+}CD8 T cells (right). [(A) to (H)] Representative experiment from two to three independent experiments. (I) Data are from two independent experiments. *P* values were determined using unpaired two-tailed Student's *t* tests. Data are expressed as means ± SEMs.

female mice (Fig. 2, H and I). Moreover, in male mice, the earliest-appearing T3.70^{+} peripheral cells were CD4/8^{dull} cells, which increasingly became CD8 T cells over the following days (Fig. 2J).

S1P1 mediates clonal eviction, which avoids thymic stress

Because immature thymocytes cannot normally exit the thymus, we investigated how immature autoreactive thymocytes were clonally evicted during negative selection. Because the sphingosine-1–phosphate receptor-1 (S1P1) is used by mature thymocytes to exit the thymus (*18*), we assessed whether S1P1 might also mediate premature clonal eviction of immature

thymocytes. CD4/8^{dull} cells in male thymi expressed S1P1, albeit in lower amounts compared with mature CD8SP thymocytes (Fig. 3A and fig. S3, A and B). Their exit from the thymus required S1P1 because in vivo S1P1 blockade by FTY720 induced short-term thymic retention, which caused T3.70^{+}CD4/8^{dull} cells to disappear from the periphery (Fig. 3B) (*19*). Moreover, conditional deletion of S1P1 in HY^{cd4}M thymocytes, which caused prolonged thymic retention, resulted in the generation of mature CD8SP thymocytes in S1P1^{cKO}.HY^{cd4}M mice that were CD24^{lo}CCR7^{hi}Qa-2^{hi} (Fig. 3, C and D, and fig. S3C). Unexpectedly, these mature CD8SP male thymocytes were autoreactive against HY self-antigens in the thymus, as they expressed interferon-γ (IFN-γ), which induced preselection CD4^{+}CD8^{+} [double-positive (DP)] thymocytes to express CD317 (an IFN-γ–responsive protein) and caused thymic stress with reduced total thymocyte numbers (Fig. 3, E to G) (*20*–*22*). Thus, S1P1-mediated clonal eviction rids the thymus of immature autoreactive cells before they become functionally competent and cause thymic stress. Moreover, the appearance of mature CD8SP thymocytes in S1P1-deficient HY^{cd4}M mice demonstrates that their absence in wild-type (WT) HY^{cd4}M mice is not a result of clonal deletion but is instead because of clonal eviction of their immature precursors from the thymus.

Gfi1 regulates S1P1 expression and clonal eviction

Mature SP thymocytes express S1P1 only after TCR signaling is disrupted because this event up-regulates the expression of the transcription factor Foxo1, which, in turn, induces the expression of *Klf2* and *S1pr1*, the gene encoding S1P1 (*23*). By contrast, immature TCR^{hi}CD4/8^{dull} thymocytes expressed S1P1 despite being TCR signaled (Fig. 3A and fig. S3, A and B). To understand the contradictory requirements for S1P1 induction in immature and mature thymocytes, we compared TCR signaling of immature CD69^{-}DP and mature CD4SP thymocytes (Fig. 3H and fig. S4A). TCR signaling up-regulated *Nr4a1* in both thymocytes but had different effects on their expression of *Foxo1*, *Klf2*, and *S1pr1* (Fig. 3H and fig. S4, A and B). Without stimulation, mature thymocytes expressed *Foxo1*, *Klf2*, and *S1pr1*, which were then down-regulated by TCR signaling (Fig. 3H and fig. S4A), whereas immature thymocytes expressed *Foxo1*, *Klf2*, and *S1pr1* only when actively TCR signaled (Fig. 3H and fig. S4, A and B). In immature thymocytes, Foxo1 deficiency abrogated *Klf2* and *S1pr1* induction, as in mature cells (fig. S4C), which indicates that Foxo1 up-regulation initiated the same transcriptional progression to S1P1 expression in immature and mature cells (*23*). Thus, the key feature distinguishing immature and mature thymocytes is that Foxo1 up-regulation in immature thymocytes is induced by strong TCR signaling, whereas Foxo1 up-regulation in mature thymocytes is induced by TCR signaling disruption.

To induce Foxo1 up-regulation in immature thymocytes, strong TCR signaling would either up-regulate a transcriptional activator or down-regulate a transcriptional repressor. Gfi1 is a known transcriptional repressor of Foxo1 (*24*), which we found to be highly expressed in immature (CD69^{-}DP) thymocytes but whose expression was reduced in more differentiated CD4/8^{dull} and CD8SP thymocytes (Fig. 3I). Gfi1 expression in immature thymocytes was down-regulated by strong TCR signaling (Fig. 3J), which indicates that strong TCR signaling of immature thymocytes induces Foxo1 up-regulation by down-regulating Gfi1 expression.

To assess the role of Gfi1 in S1P1 induction and clonal eviction, we used the lentivirus "flex-switch" system and the CD4-Cre recombinase (present in HY^{cd4} thymocytes) to induce constitutive expression of lentiviral Gfi1 or lentiviral control Thy1.1 proteins in HY^{cd4} thymocytes (fig. S4D) (*25*). Thymocytes that constitutively expressed lentiviral Gfi1 or Thy1.1 proteins also constitutively expressed hCD2 reporter proteins (fig. S4, D and E). Examination of hCD2^{+}CD4/8^{dull} thymocytes constitutively expressing Gfi1 or Thy1.1 proteins revealed that constitutive Gfi1 expression reduced *Foxo1*, *Klf2*, and *S1pr1* but did not affect PD-1 or Qa-2 (Fig. 3K and fig. S4, F and G). Gfi1-induced down-regulation of *S1pr1* resulted in thymic retention and the appearance of mature Qa-2^{+} CD8SP thymocytes (Fig. 3K and fig. S4H). Thus, it is by down-regulating Gfi1 that strong TCR signaling induces immature thymocytes to express S1P1 and undergo premature clonal eviction (fig. S5).

Peripheral CD8 T cells are self-tolerant despite expressing autoreactive TCRs

Although HY^{cd4}M mice contain peripheral CD8 T cells bearing autoreactive TCRs, these animals remain healthy throughout their lives, which suggests that their peripheral CD8 T cells are self-tolerant. Compared with female peripheral CD8 T cells, male peripheral CD8 T cells were 10-fold less reactive against antigenic HY peptide in vitro and were also less reactive to HY peptide injected in vivo as revealed by CD107a expression—a marker of CD8 T cell degranulation (*26*) (Fig. 4, A and B). Additionally, the growth of male-origin tumor cells (B16-F10) was significantly less constrained in male compared with female HY^{cd4} mice, whereas female-origin tumor cells (MC38) grew equally in both (Fig. 4C and fig. S6A). Thus, despite bearing HY-specific TCRs, peripheral CD8 T cells in male mice are hyporesponsive to HY male antigen. To determine the basis of this hyporesponsiveness, we examined surface expression of TCRs and CD8 co-receptor proteins on peripheral male CD8 T cells (Fig. 4D). Surface TCR and CD8 were both significantly reduced on male CD8 T cells, which bound only one-third the number of HY–MHC-I tetramer complexes compared with female CD8 T cells (Fig. 4D) (*27*). Thus, reduced surface TCR and co-receptor expression causes reduced binding of HY self-ligands (Fig. 4D).

We considered that male CD8 T cells may also be hyporesponsive because they have regulatory function, even though they did not express Foxp3, the master gene regulator of CD4 regulatory T cells (fig. S6B) (*28*). CD8 regulatory T cells are instead thought to be CD122^{+}Ly49^{+} and to express Helios (*29*–*31*). Peripheral CD8 T cells in HY^{cd4}M mice included a substantial number of CD122^{+}Ly49^{+} cells that expressed Helios (Fig. 4E and fig. S6C) and prevented the induction of inflammatory bowel disease in male *Rag*-deficient mice by transferred naïve CD4 T cells. Thus, clonally evicted peripheral CD8 male T cells have regulatory function (Fig. 4F).

Because peripheral CD8 male T cells contained an Ly49^{+} subset and were hyporesponsive to HY antigens, we assessed Ly49 expression on CD8SP thymocytes that arose during thymic retention and were reactive against intrathymic HY antigens (Fig. 3, E to G). Retained thymic CD8 male T cells were exclusively CD122^{+}Ly49^{-}, unlike peripheral CD8 male T cells (Fig. 4G). We therefore wondered whether their Ly49 expression was induced by encountering HY self-antigens in the periphery. Ly49 expression persisted and remained significantly more frequent after transfer of CD8 T cells into male compared with female host mice (Fig. 4H), and a substantial fraction of Ly49^{-} male cells converted into Ly49^{+} cells after transfer into male hosts, which indicates that these were not distinct subsets of autoreactive CD8 T cells (fig. S6D). Thus, immature autoreactive thymocytes evicted from the thymus encounter peripheral self-ligands to become self-tolerant CD8 T cells, some of which express Ly49.

To verify that CD8 T cell tolerance depends on clonal eviction, we compared the reactivity of CD8SP thymocytes retained in the thymus because of S1P1 deficiency versus CD8 T cells evicted into the periphery (fig. S6E). Retained CD8SP thymocytes were CD69^{hi} in vivo and further up-regulated CD69 upon exogenous in vitro stimulation with HY peptide presented by irradiated splenocytes (fig. S6E). By contrast, peripheral CD8 T cells were CD69^{lo} in vivo and hardly responded to exogenous in vitro HY peptide stimulation (fig. S6E). Thus, CD8 T cell tolerance to self-antigen depends on clonal eviction of immature autoreactive thymocytes into the

periphery before they have acquired functional competence.

Clonal eviction of normal polyclonal CD8 T cells

Finally, we wondered whether our understanding of clonal eviction in HY-specific TCR transgenic mice (schematized in fig. S7) also applied to CD8 T cells in normal polyclonal mice. We found that ~5% of MHC-I–selected TCR^{hi} thymocytes in MHC-II^- polyclonal mice were $PD\text{-}1^+$ cells (Fig. 5A). These $TCR^{hi}PD\text{-}1^+$ thymocytes were $CD4/8^{dull}$ and resembled $HY^{cd4}M$ thymocytes that had survived negative selection, whereas their $TCR^{hi}PD\text{-}1^-$ polyclonal thymocytes contained CD8SP cells and resembled $HY^{cd4}F$ thymocytes that had undergone positive selection (Fig. 5A). MHC-I–selected $TCR^{hi}PD\text{-}1^+CD4/8^{dull}$ thymocytes were present in both male and female polyclonal mice (Fig. 5B), which indicates that they had been signaled to undergo negative selection by sex-independent self-ligands. These negatively selected $PD\text{-}1^+TCR^{hi}CD4/8^{dull}$ thymocytes had down-regulated *Gfi1* and up-regulated *S1pr1* expression compared with preselection $CD69^-$DP thymocytes (Fig. 5C), which indicates that they had been strongly TCR signaled in the thymus. To determine whether these cells underwent clonal eviction into the periphery, we identified $PD\text{-}1^+$ cells among recent thymic emigrants (RTEs) by their expression of the reporter protein Rag-GFP [green fluorescent protein (GFP)] (Fig. 5D) (*32*). $PD\text{-}1^+$ RTEs were $CD4/8^{dull}$ and had the highest content of Rag-GFP, which revealed that they were the earliest RTEs in the periphery (Fig. 5E). In addition, $CD4/8^{dull}$ RTEs differentiated into mature CD8 T cells, as their mRNA expression profile was $Cd4^-Cd8^+Runx3d^+$, and they became CD8 T cells when cultured with IL-7 (Fig. 5, F and G).

MHC-I–selected peripheral CD8 T cells in polyclonal mice contained an $Ly49^+$ subset that resembled self-tolerant CD8 T cells in $HY^{cd4}M$ mice in that peripheral $Ly49^+$CD8 T cells expressed lower levels of both surface TCR and CD8 co-receptors (Fig. 5H). This resemblance suggests that peripheral $Ly49^+$ CD8 T cells may have been clonally evicted during negative selection and encountered their self-ligand during peripheral differentiation into mature CD8 T cells.

Finally, we quantified the number of CD4 and CD8 T cells bearing HY-specific autoreactive TCRs in male and female B6 mice (Fig. 5I). HY/I-A^b tetramer–binding CD4 T cells were significantly reduced in male versus female B6 mice, whereas HY/H-$2D^b$ tetramer–binding CD8 T cells were equally represented in both males and females, which indicates significant clonal deletion of HY-specific CD4 T cells but not HY-specific CD8 T cells in WT B6 mice (Fig. 5I). Additionally, HY–H-$2D^b$ tetramer–binding cells were significantly more frequent among $Ly49^+CD8^+$ T cells in male versus female mice, but this was specific for negatively selected HY–H-$2D^b$ tetramer–binding cells because it was not the case for unrelated Flu–H-$2D^b$ tetramer–binding cells, which bear TCRs specific for the influenza foreign antigen (Fig. 5I). Thus, these results further support the conclusion that $Ly49^+$ CD8 T cells are clonally evicted during negative selection and differentiated in the periphery into self-tolerant CD8 T cells.

Discussion

This study identifies premature clonal eviction from the thymus into the periphery as an important consequence of negative selection that results in extrathymic T cell differentiation of self-tolerant CD8 T cells. Premature clonal eviction is triggered by TCR-signaled down-regulation of the transcriptional repressor Gfi1 and induction of S1P1 in immature CD8 thymocytes that survive negative selection. The mechanism establishing CD8 T tolerance has remained enigmatic because functionally mature CD8 T cells bearing autoreactive TCRs are generally absent from the thymus but nevertheless appear in the periphery of both humans and mice (*33*–*40*). It has been thought that CD8 T cells with autoreactive TCRs appear in the periphery because clonal deletion in the thymus is imperfect (*36*, *40*), but this perspective does not explain how autoreactive CD8 T cells that are absent from the thymus are nevertheless undiminished in number in the periphery.

This study provides clonal eviction of immature autoreactive CD8 T cells as a solution to this enigma. Immature CD8 T cells with autoreactive TCRs are strongly TCR signaled during negative selection in the thymus to down-regulate Gfi1 and up-regulate S1P1 expression, which results in the premature eviction of immature $CD4/8^{dull}$ thymocytes from the thymus into the periphery, where they complete their differentiation into functionally mature but self-tolerant CD8 T cells, some of which become $Ly49^+$ regulatory CD8 T cells. We expect that prematurely evicted immature cells from the thymus differentiate into self-tolerant CD8 T cells in part because they are functionally immature when they first encounter peripheral self-ligands, which reduces surface expression of both TCR and CD8 co-receptor molecules, and in part because they acquire regulatory function. Thus, clonal eviction is the primary mechanism underlying CD8 T cell tolerance to self-ligands.

REFERENCES AND NOTES

1. K. M. Ashby, K. A. Hogquist, *Nat. Rev. Immunol.* 10.1038/s41577-023-00911-8 (2023).
2. T. M. McCaughtry, K. A. Hogquist, *Semin. Immunopathol.* **30**, 399–409 (2008).
3. J. W. Kappler, N. Roehm, P. Marrack, *Cell* **49**, 273–280 (1987).
4. E. Palmer, *Nat. Rev. Immunol.* **3**, 383–391 (2003).
5. E. Brugnera *et al.*, *Immunity* **13**, 59–71 (2000).
6. A. Singer, S. Adoro, J. H. Park, *Nat. Rev. Immunol.* **8**, 788–801 (2008).
7. M. Shinzawa *et al.*, *Nat. Immunol.* **23**, 731–742 (2022).
8. L. A. Pobezinsky *et al.*, *Nat. Immunol.* **13**, 569–578 (2012).
9. Y. Ishida, Y. Agata, K. Shibahara, T. Honjo, *EMBO J.* **11**, 3887–3895 (1992).
10. J. L. Riley, *Immunol. Rev.* **229**, 114–125 (2009).
11. T. A. Baldwin, M. M. Sandau, S. C. Jameson, K. A. Hogquist, *J. Exp. Med.* **202**, 111–121 (2005).
12. O. Lantz, I. Grandjean, P. Matzinger, J. P. Di Santo, *Nat. Immunol.* **1**, 54–58 (2000).
13. T. M. McCaughtry, T. A. Baldwin, M. S. Wilken, K. A. Hogquist, *J. Exp. Med.* **205**, 2575–2584 (2008).
14. J. A. Punt, B. A. Osborne, Y. Takahama, S. O. Sharrow, A. Singer, *J. Exp. Med.* **179**, 709–713 (1994).
15. L. Klein, B. Kyewski, P. M. Allen, K. A. Hogquist, *Nat. Rev. Immunol.* **14**, 377–391 (2014).
16. M. A. Zachariah, J. G. Cyster, *Science* **328**, 1129–1135 (2010).
17. R. Etzensperger *et al.*, *Nat. Immunol.* **18**, 1218–1227 (2017).
18. M. L. Allende, J. L. Dreier, S. Mandala, R. L. Proia, *J. Biol. Chem.* **279**, 15396–15401 (2004).
19. S. Mandala *et al.*, *Science* **296**, 346–349 (2002).
20. B. W. Dulken *et al.*, *Nature* **571**, 205–210 (2019).
21. Y. Kato *et al.*, *Cell. Immunol.* **177**, 103–108 (1997).
22. R. Fayad *et al.*, *Eur. J. Immunol.* **35**, 2304–2312 (2005).
23. S. M. Hedrick, R. H. Michelini, A. L. Doedens, A. W. Goldrath, E. L. Stone, *Nat. Rev. Immunol.* **12**, 649–661 (2012).
24. L. Z. Shi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E67–E74 (2017).
25. F. Schnütgen *et al.*, *Nat. Biotechnol.* **21**, 562–565 (2003).
26. M. R. Betts *et al.*, *J. Immunol. Methods* **281**, 65–78 (2003).
27. J. H. Park *et al.*, *Nat. Immunol.* **8**, 1049–1059 (2007).
28. S. Hori, T. Nomura, S. Sakaguchi, *Science* **299**, 1057–1061 (2003).
29. N. Saligrama *et al.*, *Nature* **572**, 481–487 (2019).
30. J. Li *et al.*, *Science* **376**, eabi9591 (2022).
31. H. J. Kim, B. Verbinnen, X. Tang, L. Lu, H. Cantor, *Nature* **467**, 328–332 (2010).
32. T. E. Boursalian, J. Golob, D. M. Soper, C. J. Cooper, P. J. Fink, *Nat. Immunol.* **5**, 418–425 (2004).
33. H. S. Teh, H. Kishi, B. Scott, H. Von Boehmer, *J. Exp. Med.* **169**, 795–806 (1989).
34. H. von Boehmer, J. Kirberg, B. Rocha, *J. Exp. Med.* **174**, 1001–1008 (1991).
35. S. Dhanji, S. J. Teh, D. Oble, J. J. Priatel, H. S. Teh, *Blood* **104**, 2116–2123 (2004).
36. C. Bouneaud, P. Kourilsky, P. Bousso, *Immunity* **13**, 829–840 (2000).
37. J. H. Velthuis *et al.*, *Diabetes* **59**, 1721–1730 (2010).
38. D. Kronenberg *et al.*, *Diabetes* **61**, 1752–1759 (2012).
39. Y. Maeda *et al.*, *Science* **346**, 1536–1540 (2014).
40. W. Yu *et al.*, *Immunity* **42**, 929–941 (2015).

ACKNOWLEDGMENTS

We thank J. DiCarlo, M. Y. Kimura, C. T. Mayer, J.-H. Park, D. Singer, and Y. Takahama for critically reading the manuscript; T. Baldwin and K. Hogquist for generously providing HY^{cd4} mice; M. Nussenzweig for Rag-GFP transgenic mice; A. Sharpe for B7-deficient mice; and D. Littman for $Runx3d^{yfp/yfp}$ mice. We also thank S. Schwab for providing the S1P1 staining protocol; T. Adams, K. Bisht, A. Crossman, L. Granger, W. Hajjar, and S. Sharrow for flow cytometry support; and the NIH Tetramer Core Facility for providing peptide-MHC tetramer complexes. **Funding:** This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. **Author contributions:** M.E.B. conceptualized the study, designed and performed experiments, analyzed data, and wrote the manuscript. Z.Z. and X.T. provided resources and performed experiments. A.S. conceptualized and supervised the study and wrote the manuscript. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh4124
Materials and Methods
Figs. S1 to S7
Table S1
References (*41*–*44*)
MDAR Reproducibility Checklist

Submitted 1 March 2023; resubmitted 21 June 2023
Accepted 22 September 2023
10.1126/science.adh4124

MESOSCOPIC PHYSICS

Universal chiral Luttinger liquid behavior in a graphene fractional quantum Hall point contact

Liam A. Cohen[1]†, Noah L. Samuelson[1]†, Taige Wang[2,3], Takashi Taniguchi[4], Kenji Watanabe[5], Michael P. Zaletel[2,3], Andrea F. Young[1]*

One-dimensional conductors are described by Luttinger liquid theory, which predicts a power-law suppression of the single-electron tunneling density of states at low voltages. The scaling exponent is predicted to be quantized when tunneling into a single isolated chiral edge state of the fractional quantum Hall effect. We report conductance measurements across a point contact linking integer and fractional quantum Hall edge states (at fillings 1 and $\frac{1}{3}$, respectively). At weak coupling, we observe the predicted universal quadratic scaling with temperature and voltage. At strong coupling, we demonstrate perfect Andreev reflection of fractionalized quasiparticles at the point contact. We use the strong coupling physics to realize a nearly dissipationless direct current voltage step-up transformer, whose gain arises directly from topological fractionalization of electrical charge.

The Landau theory of Fermi liquids provides a near-ubiquitous description of interacting fermion systems. One exception is provided when electrons are confined to one dimension, where arbitrarily weak interactions favor a distinct phase known as the Tomonaga-Luttinger liquid (*1*–*4*). In this phase, the low-energy collective excitations are orthogonal to the single-electron operators from which they are microscopically constructed. This "orthogonality catastrophe" manifests experimentally as a power-law suppression of the electron tunneling density of states, $N(E) \propto (E - E_F)^{(\frac{1}{g}-1)}$, at the Fermi energy E_F, even though the system remains conductive. The power law is characterized by an exponent g known as the Luttinger parameter, which depends continuously on the nature and strength of the interparticle interactions (*5*). Experimentally, Luttinger liquid behavior can manifest through a non-Ohmic current-voltage relation $I(V) \propto V^{\frac{1}{g}}$, as observed, for example, in ropes of single-walled carbon nanotubes (*6*, *7*).

Chiral Luttinger liquids

An alternative means of creating a one-dimensional (1D) wire can be found at the boundary of a topologically ordered phase, as occurs in the fractional quantum Hall (FQH) effect (*8*). Here, the right- and left- moving modes are physically separated to opposite edges of the 2D sample, resulting in a chiral Luttinger liquid in which backscattering is suppressed entirely and the Luttinger parameter g becomes quantized (*9*). In this setting, g becomes a fingerprint of the topological order of the enclosed bulk; for example $g = \frac{1}{3}$ for the Laughlin state at Landau level filling $\nu = \frac{1}{3}$ (*9*, *10*). In such a scenario, tunneling of whole electrons into the fractionalized edge of the $\nu = \frac{1}{3}$ state is predicted to exhibit a quadratic scaling of the tunneling conductance $G \equiv \frac{dI}{dV}$ with both the temperature T and bias voltage V: $G \propto T^2, V^2$. The quadratic scaling is a direct result of the quantized value of $g = \frac{1}{3}$, and a central prediction of the theory is that this behavior is universal and independent of microscopic details at sufficiently low temperatures and bias voltages (*9*).

Experimentally, one of the most successful tests of the chiral Luttinger liquid theory was obtained by studying tunneling between the edge of a 2D-electron gas (2DEG) hosting a FQH state at $\nu = \frac{1}{3}$ and a 3D electrode grown on the cleaved edge of the semiconductor wafer (*3*, *11*). Although striking power-law behavior was observed over a wide range of bias voltages and temperatures (*11*), the exponent $\frac{1}{g} \sim 2.7$ was found to vary across samples and within the $\nu = \frac{1}{3}$ plateau (*12*, *13*)—in disagreement with the predicted quantized exponent $\frac{1}{g} = 3$. This sparked a thorough discussion about the nature of bulk-boundary correspondence and whether g is indeed a distinct imprint of the topological order of the bulk (*14*–*16*). However, the possibility of edge reconstruction was soon identified as a confounding factor that could reconcile the range of observed exponents with the chiral Luttinger liquid theory (*17*–*19*).

In principle, a quantum point contact between integer and fractional quantum Hall edge states (*5*, *9*, *10*, *20*, *21*–*23*) provides a richer test bed for chiral Luttinger liquid physics. In this geometry, the collective modes of the $\nu = \frac{1}{3}$ and $\nu = 1$ edges may be modeled as chiral bosonic fields ϕ_a and ϕ_b, respectively, coupled by a single point scatterer of strength Γ. The low energy physics are described by the Lagrangian

$$\mathcal{L} = \frac{1}{4\pi} \sum_{i=a,b} \partial_x \phi_i (\partial_t - \partial_x) \phi_i + \Gamma \delta(x) \left(\psi_a^\dagger \psi_b + \psi_b^\dagger \psi_a \right) \quad (1)$$

where the operators $\psi_a = e^{i\sqrt{3}\phi_a}$ and $\psi_b = e^{i\phi_b}$ remove an electron on the $\nu = \frac{1}{3}$ and $\nu = 1$ edges, respectively. In Eq. 1, the first term describes the gapless bosonic edge modes on either side of the junction, whereas the second term describes inter-edge tunneling of electrons at the point contact. In the language of the renormalization group, the scaling dimensions of the electron operator $[\psi_b] = \frac{1}{2}$ whereas $[\psi_a] = \frac{3}{2}$ is three times larger, reflecting the topological order of the $\nu = \frac{1}{3}$ fractional quantum Hall bulk. For the corresponding 2D Euclidean action to remain dimensionless, $[\Gamma] = 1 - [\psi_a] - [\psi_b] = -1$, meaning that edge-to-edge tunneling is irrelevant and becomes increasingly less important at low energies. This leads to the conclusion that no matter how "open" the junction is made—in other words, no matter how large the bare value of Γ is— the conductance will vanish at a sufficiently low temperature and voltage bias (*24*). Near this decoupled fixed point, the conductance can be computed within perturbation theory (*3*), giving

$$G(V,T) = \frac{e^2}{2h} \left(\frac{2\pi T}{T_0} \right)^2 \left[\frac{1}{3} + \left(\frac{eV}{2\pi k_b T} \right)^2 + \ldots \right] \quad (2)$$

for small temperatures T and voltage bias V. Here, the bare tunneling strength is represented by T_0, where for weakly coupled edges $T_0 \propto \frac{1}{\Gamma}$. In this limit, the power law exponent $\left[\left(\frac{1}{g} \right) - 1 \right] = 2$ describing the T and V dependence provides a direct probe of the bulk topological order.

Quantum point contact experiments in semiconductor quantum wells have indicated power-law behavior near the decoupled fixed point (close to full pinch-off) over a limited temperature range (*25*, *26*); however, the presence of disorder at the tunnel junction often complicates the physical interpretation (*25*–*31*). In a more recent experiment (*32*) focused on the weak quasiparticle backscattering limit, the conductance showed clear characteristics of a Luttinger liquid but remained quantitatively inconsistent with the predictions of chiral Luttinger liquid theory: The measured value of the Luttinger parameter g did not align with predictions for any candidate incompressible ground states at the filling factors studied in (*31*). It appears that for both quantum point contact and cleaved-edge overgrowth experiments, nonuniversal effects arising from the detailed structure of experimentally realized edges

[1]Department of Physics, University of California at Santa Barbara, Santa Barbara, CA 93106, USA. [2]Department of Physics, University of California, Berkeley, CA 94720, USA. [3]Material Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [5]Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan.
*Corresponding author. Email: andrea@physics.ucsb.edu
†These authors contributed equally to this work.

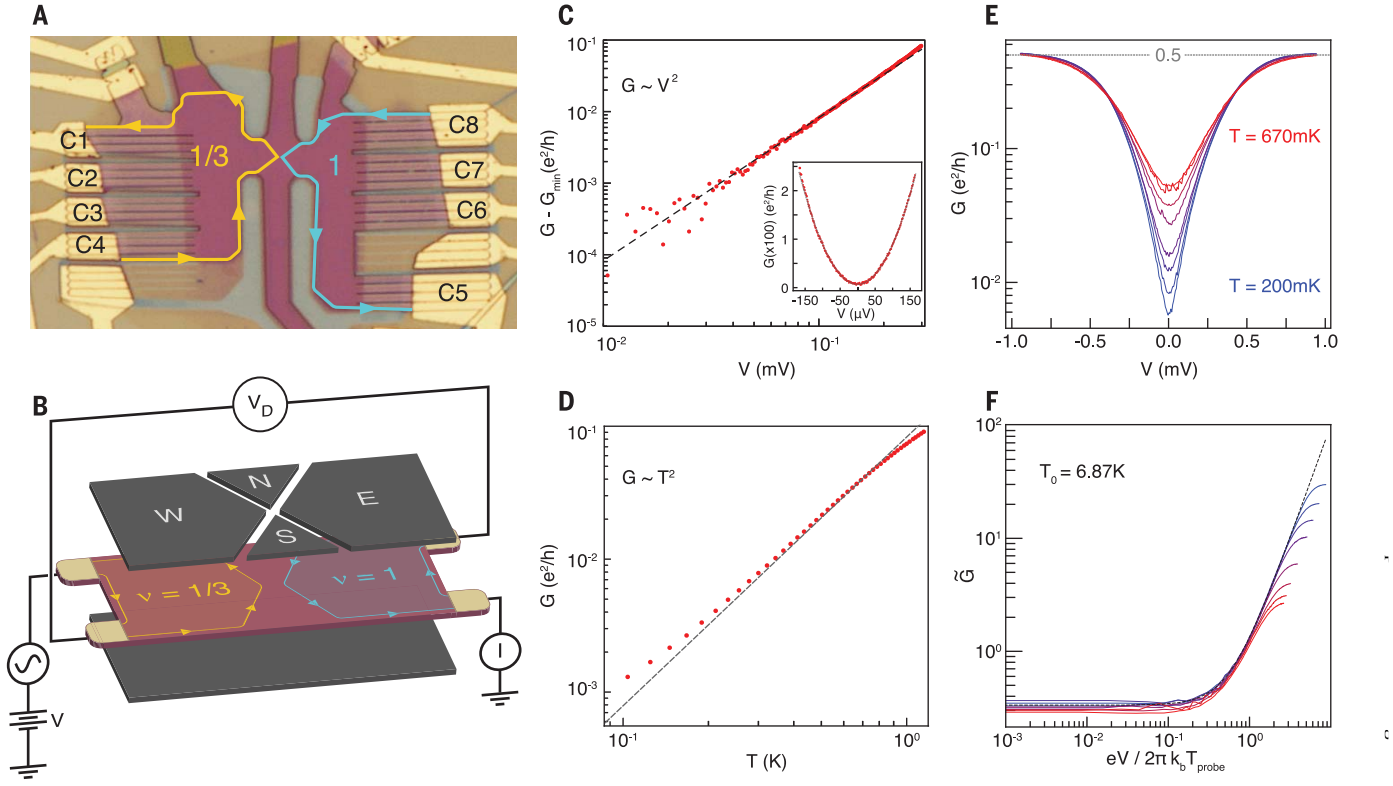


Fig. 1. Universal conductance scaling at weak coupling. (**A**) Optical micrograph of the device with schematic depiction of chiral edge states. (**B**) Device schematic showing the patterned top graphite layer, graphene monolayer, and global bottom graphite gate. The tunneling conductance across the junction is determined from the transmitted current I and diagonal voltage V_D as $G \equiv \frac{I}{V_D}$. (**C**) G measured as function of V at T_{probe} = 56 mK, with V_{NS} = −2.465 V and B = 10 T. (Inset) parabolic fit to the low-V regime, giving T_0 = 9.02 ± 0.007 K as defined in Eq. 2. The main panel shows G − G_{min}, where $G_{min} = 7.5 \times 10^{-4} \frac{e^2}{h}$ is the minimum conductance. Fitting a power law gives an exponent of 2.00 ± 0.06 (*44*), where the error represents the standard deviation in the fit parameter. (**D**) G measured at V = 0 as a function of temperature at the same gate voltages as (C). The dashed line is a plot of the conductance given by the first term of Eq. 2, using T_0 = 9.02 K. (**E**) Nonlinear differential conductance for T_{probe} = 202, 245, 290, 344, 450, 549, 618, and 666 mK at V_{NS} = −2.456 V. (**F**) The same data as in (E) after scaling G and V as described in the text. The black dashed line is $\tilde{G}$ as predicted by Eq. 2.

(*22*, *33*, *34*, *35*) push the universal scaling regime to energy scales beyond experimental reach.

Conductance scaling in graphene quantum point contact heterojunction

Recently, graphene has emerged as a useful platform for precision tests of chiral Luttinger liquid physics. Notably, the low disorder available in field effect–controlled devices make a wide range of fractional quantum Hall states accessible (*36*, *37*). However, prior studies of quantum point contacts have been limited to regimes dominated by resonant tunneling effects, owing at least in part to disorder introduced during the fabrication of local split gates (*38*–*41*). We use anodic oxidation lithography to pattern nanoscale features in graphite gates, which are then incorporated into a van der Waals heterostructure to produce a clean quantum point contact (*42*). A micrograph and schematic of our device is shown in Fig. 1, A and B. 44-nm-thick hexagonal boron nitride (hBN) dielectric layers are used as spacers between the monolayer graphene layer and the top and bottom

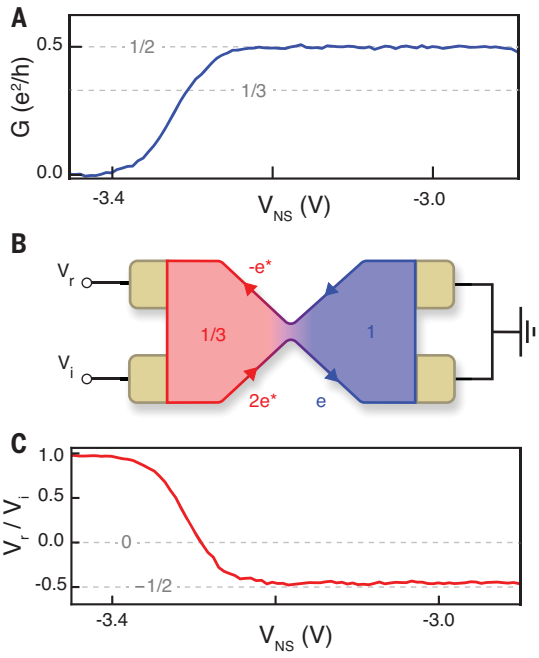


Fig. 2. Andreev-like quasiparticle scattering. (**A**) G versus V_{NS} taken at B = 9 T with a finite DC voltage bias of 145 μV. Here, V_{BG} = 2.0 V, V_E = −1.460 V, and V_W = −1.775 V to maintain ν = 1 and $\nu = \frac{1}{3}$ on the east and west sides of the junction, respectively. (**B**) Schematic representation of the Andreev scattering process for fractionally charged quasiparticles in the strong coupling limit (*23*) of a $\nu = \frac{1}{3}$ to ν = 1 point heterojunction. (**C**) Ratio of the reflected voltage V_r to the source voltage V_i versus V_{NS}; all other gate voltages are the same as in (A).

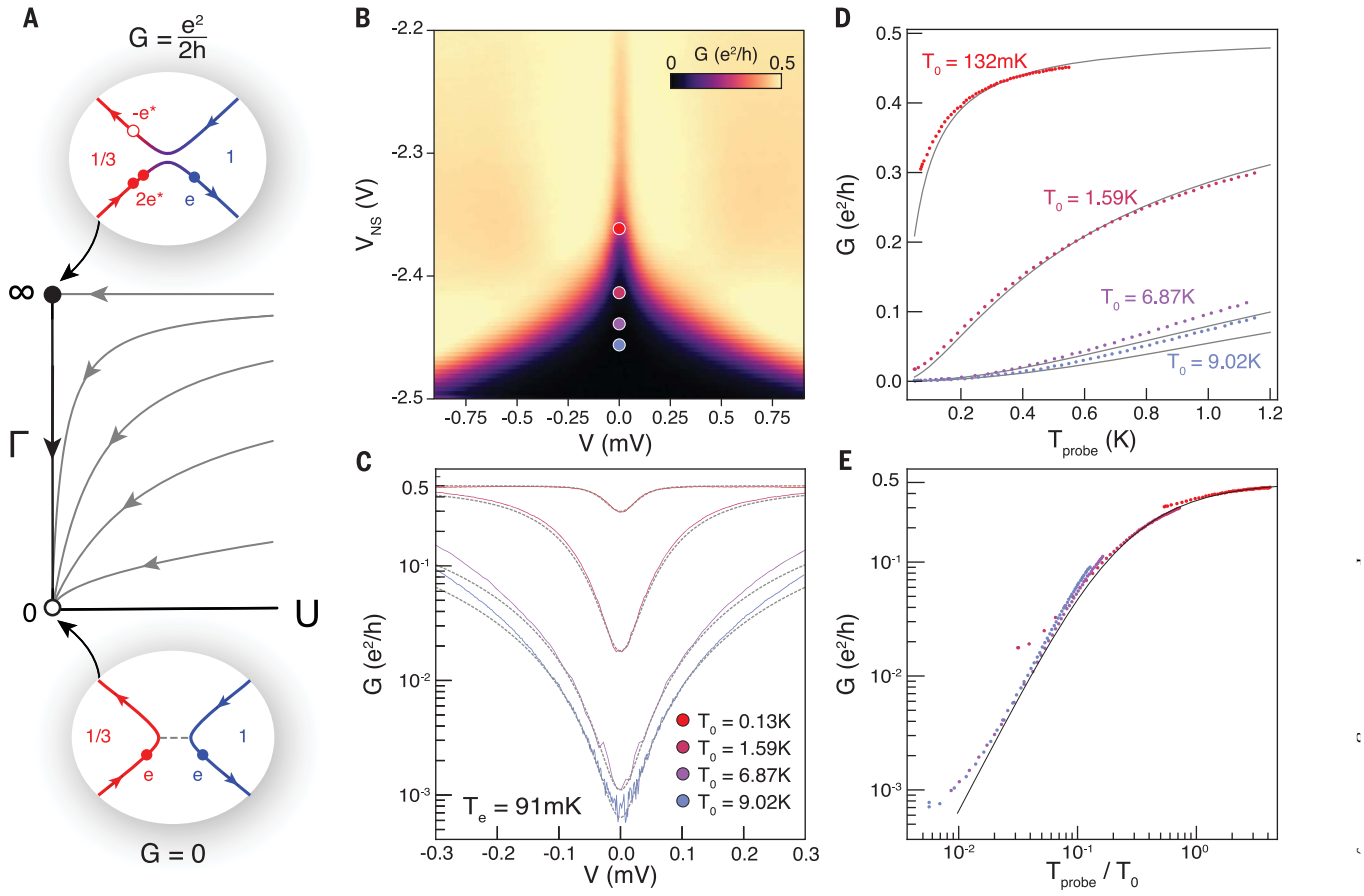


Fig. 3. Crossover from weak to strong coupling. (**A**) Schematic of renormalization group flow in the Γ–U plane, where U represents additional perturbations to Eq. 1. (**B**) G as a function of the voltage on the north and south gates, V_{NS}, and the DC voltage bias V at T_{probe} = 56 mK and B = 10 T. (**C**) Line cuts of the data in (B) at the values of V_{NS} indicated by the colored points. Black dashed lines are plotted using the value of G predicted by Eq. 3 where the parameter T_0 is extracted from the low-bias conductance. (**D**) The zero-bias conductance also scales with temperature in agreement with Eq. 3 for low energies. Although the data deviate from the model at high energies, G nevertheless exceeds $G = \frac{e^2}{3h}$ for $T \gg T_0$, indicating strong coupling at high T. (**E**) The data from (D), after scaling T by T_0. The curves collapse onto the universal scaling formula Eq. 3, shown in black.

gates. The device architecture features a four-quadrant split gate geometry (*42*) where independent voltages may be applied to the north, south, east, and west top gates (V_N, V_S, V_E, and V_W, respectively) as well as a global bottom gate (V_{BG}). With V_{BG} held constant, V_E and V_W fix the filling factors of the east and west regions to 1 and $\frac{1}{3}$, respectively, whereas $V_{NS} = V_N = V_S$ is used to create a constriction by tuning the filling of the north and south regions to $\nu \leq 0$.

The choice of V_{BG} controls the "sharpness" of the potential profile at the constriction, parameterized by the energy scale $E_V \equiv e\frac{\partial V}{\partial x}\ell_b$. Here V, which is defined as the applied potential, is purely a function of the applied gate voltages and does not take into account the finite compressibility of the 2DEG at a high magnetic field. E_V plays a key role in the physics of fractional quantum Hall edges: When the confinement energy E_V is smaller than the Coulomb energy, the edge may reconstruct (*33*, *43*), introducing additional edge modes which may push the universal tunneling behavior to experimentally inaccessible energy scales. As described in (*42*) and (*44*), the control available in our geometry allows us to access the universal regime of $E_C < E_V$ while maintaining independent control of the bulk filling factor and quantum point contact transparency.

We begin by investigating the "weak coupling" regime where $G \ll \frac{e^2}{h}$. This corresponds to the limit of $eV \ll 2\pi k_b T_0$ and $T \ll \frac{T_0}{2\pi}$ where Eq. 2 holds. We measure $G \equiv \frac{I}{V_D}$ (see Fig. 1B), which is a four-terminal measurement of the tunneling conductance (*42*, *44*). We tune T_0 through V_{NS}. Figure 1C shows G measured as a function of the direct current (DC) voltage bias V at a fixed T_{probe} = 56 mK and V_{NS} = −2.465 V. As seen in the inset, the V dependence is well fit by a parabola, with a curvature corresponding to T_0 = 9.02 K in Eq. 2. To assess the quality of the power law fit, we subtract the minimum conductance G_{min} and plot $G - G_{min}$ on a logarithmic scale (Fig. 1C). We find a simple V^2 power law over one order of magnitude in V, corresponding to two orders of magnitude in $G - G_{min}$. Specifically, we find an exponent of 2.00 ± 0.06. Although we have not studied the filling factor dependence in detail, fig. S7 shows a second measurement taken at a different magnetic field and different value of the bulk filling factor within the $\nu = \frac{1}{3}$ plateau, which yields the same exponent (*44*).

We next compare experiment with the predicted zero-bias temperature dependence, G (V = 0, T, T_0) of Eq. 2. The result for T_0 = 9.02 K is shown in Fig. 1D. In contrast to the voltage dependence, a T^2 power law is observed only for a limited range of T, between 200 and 700 mK. We attribute deviations at lower temperatures to a decoupling of the electronic temperature from T_{probe}. At high temperatures, deviations are expected as corrections to Eq. 2 become important, with significant deviations onsetting for $T > \frac{T_0}{4\pi} \approx 750mK$ (*44*). Although the power law behavior in T occurs over a limited range, the observed power law in T is consistent with the more robust power law in V,

which is a manifestation of a general scaling relation. Defining $\tilde{G} = 2G \cdot \left(\frac{T_0}{2\pi T}\right)^2 \cdot \frac{h}{e^2}$ and $x = \frac{eV}{(2\pi k_B T)}$, it follows that $\tilde{G} = \left(\frac{1}{3}\right) + x^2 + \cdots$ provides a universal low-energy collapse.

Figure 1E shows G as a function of V for several different temperatures at $V_{NS} = -2.456$ V. The same data, plotted as $\tilde{G}$ (x), are shown in Fig. 1F. In these datasets, $T_0 = 6.87$ K is determined by fitting the lowest temperature curve in Fig. 1E to Eq. 2. For low values of V where universality is expected (*44*), the curves collapse onto the universal parabolic curve expected from chiral Luttinger liquid theory.

From weak to strong coupling

As seen in Fig. 1E, G approaches $\frac{e^2}{2h}$ at high bias. One might expect that full transmission of the incoming fractional edge mode would cause the conductance to saturate at $G = \frac{e^2}{3h}$. In fact, the observed $G \approx \frac{e^2}{2h}$ can be understood from the peculiar properties of the point contact at strong coupling, defined as $eV \gg k_b T_0$ or $T \gg T_0$. The strong coupling regime is accessed most readily by lowering T_0 (i.e., increasing V_{NS}) at fixed V and T, leading to a plateau in the differential $G \approx \frac{e^2}{2h}$ (Fig. 2A).

Microscopically, the excess conductance can be understood by analogy to Andreev scattering at a metal-superconductor interface (*22*, *23*, *45*). In this picture, conduction occurs when a pair of incident quasiparticles, each with charge $e^* = -\frac{e}{3}$, is transmuted into a single electron on the ν = 1 side through the simultaneous retroreflection of a charge $\frac{e}{3}$ quasi-hole into the downstream chiral edge state (see Fig. 2B). This process leads to the observed nearly quantized increase in G. Moreover, when the Andreev process is the dominant form of charge transfer across the junction, the downstream $\nu = \frac{1}{3}$ edge hosting the retroreflected hole is expected to develop a negative chemical potential with magnitude one-half of the voltage of the incoming edge (*23*, *46*). Figure 2C shows the measured reflected voltage V_r under the same conditions as the G data in Fig. 2A. The near-quantization of both G and $\frac{V_r}{V_i}$ over the same broad range of junction transparency implies that the described Andreev process completely dominates charge transfer. This contrasts with previous experiments in semiconductor wells (*30*), where a much smaller effect was observed, likely mediated by resonant scattering.

In our discussion, we have considered only single-electron tunneling between edges, parameterized by Γ; however, additional processes such as electron co-tunneling may also contribute, which would be represented by additional operators not included in Eq. 1. Renormalization group analyses have shown (*10*) that as $\Gamma \to 0$ these terms are more irrelevant than Γ. This guarantees that regardless of microscopic details Eq. 1 becomes a good approximation to the physical system at sufficiently low energies. This is characteristic of a stable fixed point of the renormalization group and accounts for the universal scaling behavior demonstrated in Fig. 1.

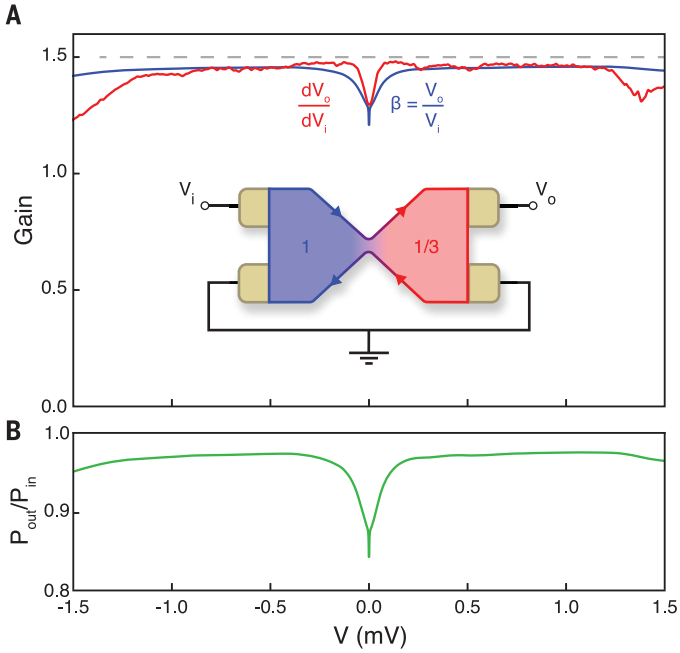


Fig. 4. Zero frequency voltage step-up transformer. (**A**) The differential gain $\frac{dV_o}{dV_i}$ and resulting integrated DC gain $\beta = \frac{dV_o}{dV_i}$, measured in the configuration shown in the inset, with B = 9 T, T_{probe} = 48 mK, $V_E = -1.460$ V, $V_W = -1.775$ V, $V_{NS} = -3.225$ V, $V_{BG} = 2.0$ V. The FQH Andreev scattering process yields an enhancement of the output voltage on the FQH side (*46*), with the DC gain predicted to reach a value of 1.5 in the dissipationless limit. Experimentally, we find a gain $\beta = 1.46$, despite the nonlinearity at low bias arising from the suppression of the Andreev scattering at low energies. (**B**) The DC power dissipation ratio, calculated from β through $P_{out}/P_{in} = \left(\frac{2\beta}{3}\right)^2 - \left(\frac{2\beta}{3}\right) + 1$, is plotted versus V, and reaches a maximum value of 97.6% (*49*).

A different result is obtained at strong coupling, formally $\Gamma \to \infty$, which represents an unstable fixed point (*23*, *47*). At this fixed point the dominant process that transfers charge across the junction is Andreev scattering. However, while other processes vanish when $\Gamma = \infty$, giving $G = \frac{e^2}{2h}$, for any finite value of Γ, G is expected to vanish at low energies as the system flows toward the stable Γ = 0 fixed point. This is shown schematically in Fig. 3A, which depicts a renormalization group flow diagram indicating the trajectory of the conductance as the energy is lowered. In this plot, the *y*-axis, Γ, represents the coefficient of the operator which transfers an electron between the two edge modes, whereas the *x*-axis represents the coefficient "U" of any operator not explicitly captured in Eq. 1. As is seen in the diagram, finite U, expected for realistic heterojunctions, would seem incompatible with approaching the strong coupling fixed point. For this reason, the strong coupling limit was previously thought to be practically inaccessible except through highly tuned resonant scattering (*46*, *48*). Although fine-tuned, this limit would arise if the QPC forms an adiabatic constriction in which approximate momentum conservation along the QPC prevents backscattering between the N and S edge (*49*).

The fact that we observe near-perfect Andreev reflection when the junction is highly transmissive suggests that the microscopics of the system approach the strong coupling fixed point with negligible contributions from the additional operators represented by U. When U = 0, Eq. 1 can be mapped to an integrable "quantum impurity model" with an exact solution for all values of Γ (*24*, *47*). The solution provides an expression for G at arbitrary V, T, and T_0,

$$G(V,T) = \frac{e^2}{2h}\int_{-\infty}^{\infty} \frac{-(2E+eV)^2}{(2E+eV)^2 + (k_b T_0)^2} f'(E)dE \quad (3)$$

where f(E) is the Fermi-Dirac function (*47*). Equation 3 provides a universal crossover function which describes the transition between weak and strong coupling, depicted as a single line along the *y*-axis (U = 0) of Fig. 3A (*22*, *23*, *47*). The integral reduces to Eq. 2 for $T_0 \to \infty$, corresponding to the weak coupling fixed point, and gives $\frac{e^2}{2h}$ for $T_0 = 0$ corresponding to strong coupling.

Figure 3B shows G as a function of V and V_{NS} at a probe temperature $T_{probe} = 56$ mK. Throughout the plotted range, the high-V conductance saturates to approximately $\frac{e^2}{2h}$ as expected for the strong coupling fixed point. Meanwhile, a zero-bias dip remains visible for all V_{NS}, indicative of the instability of the point contact to edge decoupling at low energies. To quantify how closely the quantum impurity model describes our system, we compare Eq. 3 with our experimental data as a function of V and T, and T_0. In Fig. 3C, we plot four curves extracted for different values of V_{NS}. For each curve, T_0 is determined from a fit to the low bias behavior with an appropriate low-energy expansion of Eq. 3 (*44*). For the largest fit values of T_0, the residual value of G when V = 0

can then be used as a primary thermometer—in effect allowing us to correct for a possible lack of equilibration between the electron temperature and the probe thermometer. Using this method, we find $T_{electron}$ = 91 mK for the T_0 = 9.02 K trace, in contrast to the measured probe temperature of 56 mK. Taking 91 mK as the electron temperature for the remaining data sets in Fig. 3C, Eq. 3 may then be used to generate V-dependent curves interpolating between weak and strong coupling. These curves are overlaid in black on the experimental data in Fig. 3C.

Figure 3D shows G (V = 0, T) plotted as a function of the probe temperature for the same values of V_{NS} as in Fig. 3C, along with the results of Eq. 3 for the corresponding values of T_0. Although Eq. 3 does not provide a simple scaling between temperature and voltage as is available at the weak coupling fixed point, for V = 0 the conductance can be written as a function of the scaled temperature, $\frac{T}{T_0}$. Figure 3E shows the V = 0 conductance plotted as a function of $\frac{T}{T_0}$. The four data sets shown in unscaled form in Fig. 3E collapse onto different parts of the universal crossover function between weak and strong coupling. Note that for Fig. 3, D and E, the temperature is measured on the probe and no corrections are made for disequilibrium between $T_{electron}$ and T_{probe}, leading to systematic deviations between experiment and theory at the lowest temperatures. The collapse of the experimental data onto a single universal curve over nearly two orders of magnitude in T_0 strongly supports the conclusion that our quantum point contact realizes the exactly solvable Lagrangian of Eq. 1 to a high degree of accuracy, and in particular, that tunneling predominantly occurs through a single point at the QPC rather than by multiple disorder-induced scattering centers.

Near-dissipationless step-up transformer

At the strong coupling fixed point the quantum point contact acts as a nearly dissipationless splitter, partitioning current injected on the ν = 1 edge equally between the downstream $\nu = \frac{1}{3}$ and the upstream ν = 1 edge states (*23*). For Γ = ∞, the partitioning happens with unit probability. In this limit, no entropy is generated and dissipation does not occur, leading to unity power efficiency (*46*). This contrasts with the more conventional case of partial transmission of edge modes at a quantum point contact, where fluctuations arising from partition noise lead to dissipation. The dissipationless current splitting property of the strong coupling fixed point allows us to operate our device as a voltage step-up transformer (*23*, *46*). Figure 4 shows the differential gain, $\frac{dV_o}{dV_i}$, where V_i is the input voltage applied to the upstream ν = 1 edge state and V_o is the output voltage measured on the downstream $\nu = \frac{1}{3}$ edge state with all other contacts grounded. The differential gain is very close to the theoretically expected value of $\frac{3}{2}$ (*46*). Because it is built on a purely zero frequency effect, the transformer gain remains almost the same even in the "direct current" (DC) limit, as can be seen in the behavior of the integrated gain, $\frac{V_o}{V_i}$, in comparison to its differential counterpart. The exceptionally high power efficiency, which peaks at 97.6%, is a testament to how closely the experiment realizes the strong-coupling fixed point and contrasts favorably with zero frequency voltage amplification based on superconductors (*50*, *51*) and bilayer quantum Hall systems (*49*, *51*).

Discussion and outlook

Our observation of universal chiral Luttinger liquid physics at both weak and strong coupling directly paves the way for experiments on two dimensional systems where mesoscopic electrostatic control plays a key role in addressing unanswered questions about strong correlations, topological order, and quantum statistics. Examples include even denominator fractional quantum Hall states observed in mono- (*52*, *53*) and bilayer (*38*, *54*, *55*) graphene, where taming edge reconstruction in a quantum point contact may allow for unambiguous experimental constraints on the ground state topological order (*45*, *56*–*59*). In addition, the flexibility inherent in anodic oxidation prepatterning will allow independent tuning of quasiparticle number, edge sharpness, and quantum point contact transparency in interferometer geometries where direct access to quantum statistics is possible (*60*, *61*). Finally, gate-defined point contacts may allow for precision measurements of order parameters in the recently discovered crystalline graphene superconductors (*62*–*64*).

REFERENCES AND NOTES

1. J. M. Luttinger, *J. Math. Phys.* **4**, 1154–1162 (1963).
2. S.-I. Tomonaga, *Prog. Theor. Phys.* **5**, 544 (1950).
3. A. M. Chang, *Rev. Mod. Phys.* **75**, 1449–1505 (2003).
4. T. Giamarchi, T. Nattermann, P. L. Doussal, *Quantum Physics in One Dimension* (Clarendon Press, 2003).
5. M. P. A. Fisher, L. I. Glazman, *Mesoscopic Electron Transport*, L. L. Sohn, L. P. Kouwenhoven, G. Schön, eds., NATO ASI Series (Springer, 1997), pp. 331–373.
6. M. Bockrath *et al*., *Nature* **397**, 598–601 (1999).
7. H. Ishii *et al*., *Nature* **426**, 540–544 (2003).
8. B. I. Halperin, *Phys. Rev. B Condens. Matter* **25**, 2185–2190 (1982).
9. X. G. Wen, *Phys. Rev. B Condens. Matter* **41**, 12838–12844 (1990).
10. C. L. Kane, M. P. A. Fisher, *Perspectives in Quantum Hall Effects* (Wiley, 1996), pp. 109–159.
11. A. M. Chang, L. N. Pfeiffer, K. W. West, *Phys. Rev. Lett.* **77**, 2538–2541 (1996).
12. M. Grayson, D. C. Tsui, L. N. Pfeiffer, K. W. West, A. M. Chang, *Phys. Rev. Lett.* **80**, 1062–1065 (1998).
13. A. M. Chang, M. K. Wu, C. C. Chi, L. N. Pfeiffer, K. W. West, *Phys. Rev. Lett.* **86**, 143–146 (2001).
14. V. J. Goldman, E. V. Tsiper, *Phys. Rev. Lett.* **86**, 5841 (2001).
15. S. S. Mandal, J. K. Jain, *Phys. Rev. Lett.* **89**, 096801 (2002).
16. S. Jolad, J. K. Jain, *Phys. Rev. Lett.* **102**, 116801 (2009).
17. X. Wan, F. Evers, E. H. Rezayi, *Phys. Rev. Lett.* **94**, 166804 (2005).
18. Z.-X. Hu, R. N. Bhatt, X. Wan, K. Yang, *Phys. Rev. Lett.* **107**, 236806 (2011).
19. D. Varjas, M. P. Zaletel, J. E. Moore, *Phys. Rev. B Condens. Matter Mater. Phys.* **88**, 155314 (2013).
20. D.-H. Lee, X.-G. Wen, *Phys. Rev. Lett.* **66**, 1765–1768 (1991).
21. I. Safi, *Ann. Phys.* **22**, 463–679 (1997).
22. C. C. Chamon, E. Fradkin, *Phys. Rev. B Condens. Matter* **56**, 2012 (1997).
23. N. P. Sandler, C. C. Chamon, E. Fradkin, *Phys. Rev. B Condens. Matter Mater. Phys.* **57**, 12324–12332 (1998).
24. C. L. Kane, M. P. A. Fisher, *Phys. Rev. B Condens. Matter* **46**, 15233–15262 (1992).
25. F. P. Milliken, C. P. Umbach, R. A. Webb, *Solid State Commun.* **97**, 309–313 (1996).
26. I. Maasilta, V. Goldman, *Phys. Rev. B Condens. Matter* **55**, 4081–4084 (1997).
27. S. Roddaro *et al*., *Phys. Rev. Lett.* **90**, 046805 (2003).
28. I. P. Radu *et al*., *Science* **320**, 899–902 (2008).
29. J. B. Miller *et al*., *Nat. Phys.* **3**, 561–565 (2007).
30. M. Hashisaka *et al*., *Nat. Commun.* **12**, 2794 (2021).
31. S. Baer *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **89**, 085424 (2014).
32. S. Baer *et al*., *Phys. Rev. B Condens. Matter Mater. Phys.* **90**, 075403 (2014).
33. Cd. Chamon, X. G. Wen, *Phys. Rev. B Condens. Matter* **49**, 8227–8241 (1994).
34. X. Wan, E. H. Rezayi, K. Yang, *Phys. Rev. B Condens. Matter Mater. Phys.* **68**, 125307 (2003).
35. K. Yang, *Phys. Rev. Lett.* **91**, 25 (2003).
36. A. A. Zibrov *et al*., *Nature* **549**, 360–364 (2017).
37. C. Dean, P. Kim, J. I. A. Li, A. Young, *Fractional Quantum Hall Effects: New Developments* (World Scientific, 2020), pp. 317–375.
38. K. Zimmermann *et al*., *Nat. Commun.* **8**, 14983 (2017).
39. H. Overweg *et al*., *Nano Lett.* **18**, 553–559 (2018).
40. C. Déprez *et al*., *Nat. Nanotechnol.* **16**, 555–562 (2021).
41. Y. Ronen *et al*., *Nat. Nanotechnol.* **16**, 563–569 (2021).
42. L. A. Cohen *et al*., Tunable fractional quantum Hall point contacts in graphene via local anodic oxidation of graphite gates. arXiv:0706.1234 [cond-mat.mes-hall] (2022).
43. X. Wan, K. Yang, E. H. Rezayi, *Phys. Rev. Lett.* **88**, 056802 (2002).
44. Materials and methods can be found in the supplementary materials.
45. M. Milovanović, N. Read, *Phys. Rev. B Condens. Matter* **53**, 13559–13582 (1996).
46. D. B. Chklovskii, B. I. Halperin, *Phys. Rev. B Condens. Matter Mater. Phys.* **57**, 3781–3784 (1998).
47. P. Fendley, H. Saleur, N. P. Warner, *Nucl. Phys. B* **430**, 577–596 (1994).
48. C. L. Kane, Resonant Tunneling between quantum Hall states at filling \nu = 1 and \nu = 1/3. arXiv:9809020 [cond-mat.mes-hall] (1998).
49. B. I. Halperin, A. Stern, S. M. Girvin, *Phys. Rev. B Condens. Matter Mater. Phys.* **67**, 235313 (2003).
50. I. Giaever, *IEEE Spectr.* **3**, 117–122 (1966).
51. S. M. Girvin, D. C. Transformer, D. C. Josephson, *Phys. Scr.* **2002**, 112 (2002).
52. A. A. Zibrov *et al*., *Nat. Phys.* **14**, 930–935 (2018).
53. Y. Kim *et al*., *Nat. Phys.* **15**, 154–158 (2019).
54. D.-K. Ki, V. I. Fal'ko, D. A. Abanin, A. F. Morpurgo, *Nano Lett.* **14**, 2135–2139 (2014).
55. J. Li *et al*., *Science* **358**, 648–652 (2017).
56. X.-G. Wen, *Phys. Rev. Lett.* **70**, 355–358 (1993).
57. P. Fendley, M. P. A. Fisher, C. Nayak, *Phys. Rev. Lett.* **97**, 036801 (2006).
58. R. Ohashi, R. Nakai, T. Yokoyama, Y. Tanaka, K. Nomura, *J. Phys. Soc. Jpn.* **91**, 123703 (2022).
59. W. Bishara, C. Nayak, *Phys. Rev. B Condens. Matter Mater. Phys.* **77**, 165302 (2008).
60. C. Nayak, S. H. Simon, A. Stern, M. Freedman, S. Das Sarma, *Rev. Mod. Phys.* **80**, 1083–1159 (2008).
61. J. Nakamura *et al*., *Nat. Phys.* **15**, 563–569 (2019).
62. H. Zhou, T. Xie, T. Taniguchi, K. Watanabe, A. F. Young, *Nature* **598**, 434–438 (2021).
63. H. Zhou *et al*., *Science* **375**, 774–778 (2022).
64. Y. Zhang *et al*., *Nature* **613**, 268–273 (2023).
65. L. A. Cohen *et al*., Universal chiral Luttinger liquid behavior in a graphene fractional quantum Hall point contact, Version 1.0, Zenodo (2022); https://doi.org/10.5281/zenodo.7369945.

ACKNOWLEDGMENTS

The authors acknowledge discussions with B. L. Halperin, C. Nayak, and C. Kane, and A. Assouline and A. M. Potts for helpful comments on the manuscript. **Funding:** Work at UCSB was primarily supported by the Air Force Office of Scientific Research under award FA9560-20-1-0208 and by the Gordon and Betty Moore Foundation EPIQS program under award GBMF9471. L.C. and N.S. received additional support from the Army Research Office under award W911NF20-1-0082. T.W. and M.Z. were supported by the Director, Office of Science, Office of Basic Energy Sciences,

Materials Sciences and Engineering Division of the US Department of Energy under contract no. DE-AC02-05-CH11231 (van der Waals heterostructures program, KCWF16). M.Z. received additional support from the Army Research Office through the MURI program (grant number W911NF-17-1-0323) K.W. and T.T. acknowledge support from JSPS KAKENHI (Grant Numbers 19H05790, 20H00354 and 21H05233). **Author contributions:** L.A.C. and N.L.S. fabricated the device and performed the measurements. L.A.C., N.L.S., M.P.Z., T.W., and A.F.Y. analyzed the data and wrote the paper. T.T. and K.W. synthesized the hBN crystals.

Competing interests: A.F.Y. is a board member of sp2 quantum. L.A.C. and A.F.Y. are inventors on the patent application PCT/US23/14059, which is related to this work. **Data and materials availability:** All of the experimental data presented and code for generating the figures from the datasets are available in a public Zenodo repository (*65*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf9728
Materials and Methods
Supplementary Text
Figs. S1 to S10
Reference (*66*)

Submitted 29 November 2022; resubmitted 17 December 2022
Accepted 29 September 2023
10.1126/science.adf9728

DIOXYGEN ACTIVATION

Reactive high-spin iron(IV)-oxo sites through dioxygen activation in a metal–organic framework

Kaipeng Hou[1,2]†, Jonas Börgel[1,2]†, Henry Z. H. Jiang[1], Daniel J. SantaLucia[3,4], Hyunchul Kwon[1], Hao Zhuang[5,6], Khetpakorn Chakarawet[7]‡, Rachel C. Rohde[1], Jordan W. Taylor[1], Chaochao Dun[8], Maria V. Paley[1,2]§, Ari B. Turkiewicz[1], Jesse G. Park[1]¶, Haiyan Mao[5], Ziting Zhu[2,6], E. Ercan Alp[9], Jiyong Zhao[9], Michael Y. Hu[9], Barbara Lavina[9,10], Sergey Peredkov[3], Xudong Lv[1], Julia Oktawiec[11], Katie R. Meihaus[1], Dimitrios A. Pantazis[4], Marco Vandone[12], Valentina Colombo[12,13], Eckhard Bill[3], Jeffrey J. Urban[2,8], R. David Britt[7,14], Fernande Grandjean[15], Gary J. Long[15], Serena DeBeer[3], Frank Neese[4], Jeffrey A. Reimer[2,5], Jeffrey R. Long[1,2,5]*

In nature, nonheme iron enzymes use dioxygen to generate high-spin iron(IV)=O species for a variety of oxygenation reactions. Although synthetic chemists have long sought to mimic this reactivity, the enzyme-like activation of O_2 to form high-spin iron(IV) = O species remains an unrealized goal. Here, we report a metal–organic framework featuring iron(II) sites with a local structure similar to that in α-ketoglutarate-dependent dioxygenases. The framework reacts with O_2 at low temperatures to form high-spin iron(IV) = O species that are characterized using in situ diffuse reflectance infrared Fourier transform, in situ and variable-field Mössbauer, Fe Kβ x-ray emission, and nuclear resonance vibrational spectroscopies. In the presence of O_2, the framework is competent for catalytic oxygenation of cyclohexane and the stoichiometric conversion of ethane to ethanol.

The development of catalysts for the selective oxygenation of light hydrocarbons using O_2 remains a formidable but important challenge in the global effort to develop green technologies for the valorization of natural gas components (*1*–*3*). Nature has developed mononuclear nonheme iron metalloenzymes that use O_2 for C–H oxygenation chemistry, such as the ubiquitous α-ketoglutarate-dependent dioxygenases (*4*). One well-studied enzyme in this class is taurine–α-ketoglutarate dioxygenase (TauD), which oxygenates one of the C–H bonds of taurine alpha to the sulfonate group (*5*). Key to the reactivity of TauD and its family of dioxygenases is a high-spin (S = 2) Fe(IV) = O intermediate, which is formed following oxidation of iron(II) with O_2 coupled with oxidation and decarboxylation of the α-ketoglutarate co-substrate (Fig. 1A) (*6*, *7*). Over the last several decades, significant research effort has been devoted to the design and study of iron (IV)–oxo species in molecular (*8*–*17*) and iron–zeolite (*18*, *19*) model systems in order to better understand and mimic their reactivity in biological systems. However, most examples studied to date feature an intermediate spin ground state (S = 1), and only a small number of these are generated using dioxygen in solution (*10*–*13*). High-spin Fe(IV) = O species have been accessed with oxidants such as trimethylammonium-*N*-oxide, hypervalent iodine reagents, and nitrous oxide (*14*, *15*, *18*), as well as in the presence of O_2 with light irradiation to cleave the O–O bond (*17*). However, the use of O_2 alone for the generation of high-spin Fe(IV)=O species, in a manner akin to metalloenzyme reactivity, has yet to be achieved in any synthetic system (*16*, *19*, *20*).

Metal–organic frameworks have received increasing attention in recent years as attractive systems for studying biomimetic chemistry (*21*–*23*). These porous, crystalline solids are constructed from metal nodes and organic linkers and exhibit chemical and structural tunability unmatched in other porous materials (*24*, *25*). As such, metal–organic frameworks offer the opportunity to explore O_2 activation in solid–gas reactions, while the immobilization of metal sites in the lattice may serve to prevent the decomposition of reactive species through dimerization or intramolecular ligand oxidation pathways available to molecular compounds (*9*, *26*, *27*). However, reported mimics of nonheme iron enzymes in metal–organic frameworks are scarce (*28*). In this context, framework $Fe_{1.5}Zn_{3.5}Cl_4(btdd)_3$ (H_2btdd = (1*H*-1,2,3-triazolo [4,5-*b*],[4′,5′-*i*])dibenzo[1,4] dioxin) (*29*) stands out as a suitable biomimetic platform (*22*). This material is synthesized through postsynthetic iron(II) exchange in $Zn_5Cl_4(btdd)_3$ (MFU-4*l*) (*30*) and features pseudo-tetrahedral iron(II) sites with tris(triazolate) coordination reminiscent of the binding of the 2-His-1-carboxylate facial triad in TauD. Herein, we report the frameworks $Fe_xZn_{5-x}(prv)_4(btdd)_3$ (x = 1 or 1.8; Hprv = pyruvic acid) and $FeZn_4(moba)_4(btdd)_3$ (Hmoba = 3,3-dimethyl-2-oxobutanoic acid), which react with O_2 to generate high-spin (S = 2) Fe(IV) = O species that are reactive toward hydrocarbon oxygenation (Fig. 1, B and C).

Synthesis and characterization of $Fe_xZn_{5-x}(prv)_4(btdd)_3$

The frameworks $Fe_xZn_{5-x}(prv)_4(btdd)_3$ were prepared through post-synthetic cation exchange in $Zn_5Cl_4(btdd)_3$ using ferrous chloride and subsequent ligand exchange with pyruvate (see supplementary materials). Energy-dispersive x-ray (EDX) spectroscopy revealed that the iron sites in $Fe_xZn_{5-x}Cl_4(btdd)_3$ are homogeneously distributed within the materials (figs. S1 and S2), and inductively coupled plasma optical emission spectroscopy (ICP-OES) confirmed the extent of iron substitution. Quantitative exchange of chloride for pyruvate in $Fe_xZn_{5-x}(prv)_4(btdd)_3$ is supported by elemental analysis (see section 1.3 of the SM).

Powder x-ray diffraction analysis confirmed that the $Fe_xZn_{5-x}(prv)_4(btdd)_3$ materials are crystalline solids and isostructural to the parent MFU-4*l* framework (fig. S3), and N_2 adsorption data obtained at 77 K revealed high Brunauer–Emmett–Teller (BET) surface areas of 2130 ± 12 and 2090 ± 15 m^2/g for x = 1 and 1.8, respectively (figs. S4 and S5). The 5 K Mössbauer spectrum of desolvated $FeZn_4(prv)_4(btdd)_3$ features a major doublet [area of 84.2(8)%] with an isomer shift (δ) of 1.061(1) mm/s and a quadrupole splitting ($|\Delta E_Q|$) of 2.567(1) mm/s, indicative of high-spin, five-coordinate iron(II) (fig. S18) (*31*). Two minor doublets with slightly different Mössbauer parameters likely correspond to six-coordinate iron(II) sites featuring residual coordinating synthesis solvent (table S5). The magnetic circular dichroism spectrum of

[1]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Max Planck Institute for Chemical Energy Conversion, D-45470 Mülheim an der Ruhr, Germany. [4]Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany. [5]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, USA. [6]Department of Materials Science and Engineering, University of California, Berkeley, CA 94720, USA. [7]Department of Chemistry, University of California, Davis, CA 95616, USA. [8]The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [9]Advanced Photon Source, Argonne National Laboratory, Lemont, IL 60439, USA. [10]Center for Advanced Radiation Sources, The University of Chicago, Chicago, IL 60637, USA. [11]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [12]Department of Chemistry, University of Milan, 20133 Milan, Italy. [13]Consorzio Interuniversitario Nazionale per la Scienza e Tecnologia dei Materiali (INSTM), UdR Milano, Via Golgi 19, 20133 Milano, Italy. [14]Miller Institute for Basic Research in Science, University of California, Berkeley CA 94720, USA [15]Department of Chemistry, Missouri University of Science and Technology, University of Missouri, Rolla, MO 65409, USA.
*Corresponding author. Email: jrlong@berkeley.edu
†These authors contributed equally to this work.
‡Present address: Department of Chemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand.
§Present address: ChemFinity Technologies, Inc., 710 Parkside Ave, Brooklyn, NY 11226.
¶Present address: Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

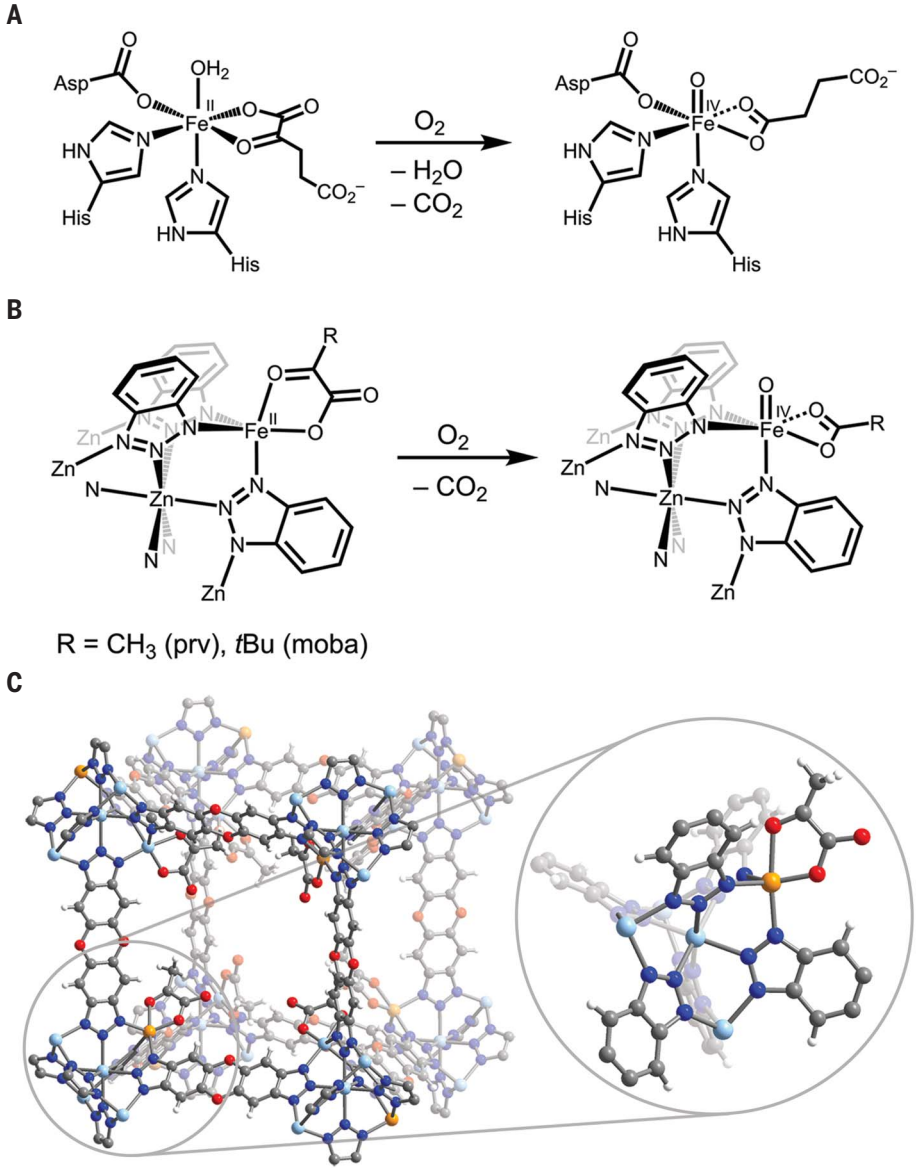


Fig. 1. Design of a metal–organic framework mimic of taurine–α-ketoglutarate dioxygenase (TauD). (**A**) Illustration of the local structure of the mononuclear nonheme iron(II) sites in TauD and generation of the reactive high-spin Fe(IV)=O species (TauD-*J*) through oxidation with O_2 coupled with decarboxylation of the α-ketoglutarate co-substrate. (**B**) Illustration of the local coordination environment of the iron(II) sites in $FeZn_4(prv)_4(btdd)_3$ (R = CH_3) or $FeZn_4(moba)_4(btdd)_3$ (R = *t*Bu) and observed reactivity with O_2 at low temperatures to form an Fe(IV)=O species coordinated by acetate or pivalate formed through the decarboxylation of pyruvate (prv) or 3,3-dimethyl-2-oxobutyrate (moba), respectively. (**C**) (Left) Illustration of a cubic pore within $FeZn_4(prv)_4(btdd)_3$ derived using single-crystal x-ray diffraction data obtained for $Zn_5(prv)_4(btdd)_3$ (fig. S15B) and (right) expanded view of the truncated structure of a cluster node of the framework showing the nature of the pyruvate coordination, as supported by Mössbauer, magnetic circular dichroism, and nuclear resonance vibrational spectroscopies. Light blue, orange, red, blue, gray, and white spheres represent Zn, Fe, O, N, C, and H atoms, respectively.

$FeZn_4(prv)_4(btdd)_3$ collected at 5 K under a field of 7 T features d-d transitions around 6000 and 14,000 cm^{-1}, which are consistent with $S = 2$ iron(II) centers in a five-coordinate ligand field (fig. S8). Continuous-wave parallel mode electron paramagnetic resonance (EPR) spectroscopy and dc magnetic susceptibility data collected for $Fe_xZn_{5-x}(prv)_4(btdd)_3$ further support the assignment of $S = 2$ for the iron(II) sites (figs. S9 to S11).

We turned to solid-state ^{1}H NMR spectroscopy to gain more insight into the possible binding modes of the pyruvate ligands in $FeZn_4(prv)_4(btdd)_3$. The spectrum exhibits broad features owing to the paramagnetic sites, which preclude the extraction of meaningful information (fig. S12). As such, we prepared a crystalline powder sample of $Zn_5(prv)_4(btdd)_3$ featuring pyruvate labeled with ^{13}C at the methyl carbon. The solid-state ^{1}H NMR spectrum features resonances from the protons of the $btdd^{2-}$ linker and the pyruvate methyl group in a ~1:1 ratio, supporting quantitative ligand exchange (fig. S13). We were also able to prepare single crystals of $Zn_5(prv)_4(btdd)_3$ (see sections 1.6 and 1.7 of the SM), and analysis through single-crystal x-ray diffraction revealed that pyruvate coordinates to the peripheral zinc(II) centers in a bidentate fashion (figs. S14B and S15B). It was not possible to isolate single crystals of $FeZn_4(prv)_4(btdd)_3$ through single-crystal-to-single-crystal exchange starting from $Zn_5Cl_4(btdd)_3$. However, the powder x-ray diffraction patterns of $Fe_xZn_{5-x}(prv)_4(btdd)_3$ are consistent with the simulated pattern generated for $Zn_5(prv)_4(btdd)_3$ from the single-crystal structure (fig. S3), which may indicate that the coordination mode of the pyruvate ligand is similar in the three frameworks.

In situ DRIFTS analysis of the reactivity of $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ with O_2

Reactivity between $Fe_xZn_{5-x}(prv)_4(btdd)_3$ and O_2 was examined with variable temperature in situ diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) (fig. S16). For this purpose, we used $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ with the goal of maximizing the resulting spectral signal whereas $FeZn_4(prv)_4(btdd)_3$ was used for the remaining spectroscopic analyses described below. Following dosing of a sample of desolvated $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ with O_2 at 100 K, a new absorption band gradually appeared in the DRIFTS spectrum at 2341 cm^{-1}, which we assign as the asymmetric C=O stretch of physisorbed CO_2 formed from decarboxylation of pyruvate (Fig. 2A, solid lines). The intensity of this band increased as the temperature was increased to 150 and 200 K; above 250 K the band disappeared, consistent with CO_2 desorption from the framework. To verify that the detected CO_2 was derived from pyruvate and not from O_2, we carried out an analogous in situ experiment using $Fe_{1.8}Zn_{3.2}(1\text{-}^{13}C\text{-}prv)_4(btdd)_3$ ($1\text{-}^{13}C\text{-}prv^-$ = pyruvate labeled with ^{13}C on the carboxylate carbon atom) and $^{18}O_2$ (Fig. 2A, dashed lines). Upon dosing with $^{18}O_2$ at 100 K, a new stretch appeared at 2275 cm^{-1}, consistent with formation of the isotopologue $^{13}CO_2$ and not $^{13}C^{18}O_2$, confirming that the oxygen atoms do not originate from dioxygen. The $^{13}CO_2$ stretch grew in intensity with heating to 200 K and disappeared at higher temperatures. A slight deviation of the experimental CO_2 stretching frequencies reported herein from the values associated with gas-phase CO_2 (2349 and 2284 cm^{-1} for CO_2 and $^{13}CO_2$, respectively)

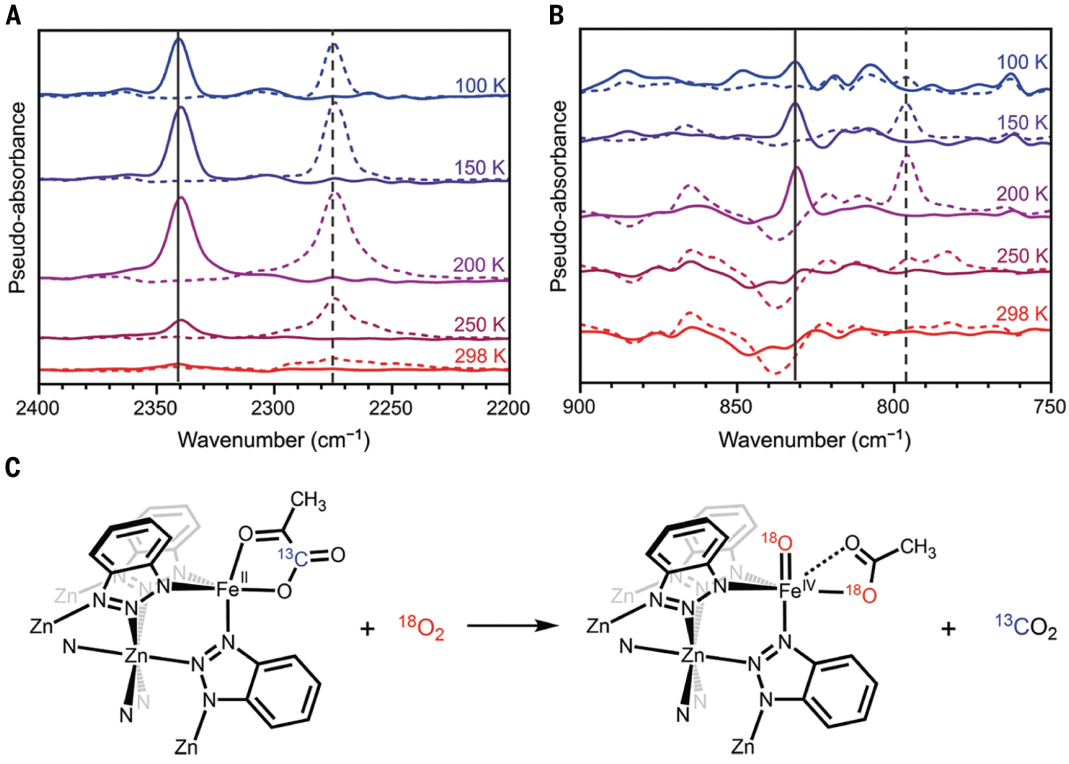


Fig. 2. Investigation of reactivity between $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ and O_2 using variable-temperature in situ DRIFTS. (**A**) Spectra obtained after dosing $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ with 20 mbar of O_2 at 100 K and gradually warming to 298 K (solid lines). The peak at 2341 cm^{-1} corresponds to CO_2 formed from the decarboxylation of pyruvate. This peak shifts to ~2275 cm^{-1} when $Fe_{1.8}Zn_{3.2}(1\text{-}^{13}C\text{-}prv)_4(btdd)_3$ is dosed with $^{18}O_2$ under the same conditions (dashed lines), consistent with formation of the isotopologue $^{13}CO_2$. (**B**) Spectra obtained as described in (A) showing the signature peaks for the iron(IV)=O species formed from the reaction between $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ and O_2 (831 cm^{-1}) and $Fe_{1.8}Zn_{3.2}(1\text{-}^{13}C\text{-}prv)_4(btdd)_3$ and $^{18}O_2$ (796 cm^{-1}). All data shown correspond to difference spectra obtained using the desolvated iron(II) frameworks as the background. Minor differences in the intensities of the absorption bands for the natural abundance and heavier isotopologue samples are likely due to differences in powder sample mass, sample distribution in the background matrix, and slight positioning differences in the infrared beam. (**C**) The reaction probed by DRIFTS, wherein for simplicity, $FeZn_4(1\text{-}^{13}C\text{-}prv)_4(btdd)_3$ is shown to react with $^{18}O_2$.

can be ascribed to adsorption of the CO_2 within the pores of the framework at low temperatures. Notably, in situ powder x-ray diffraction data collected for $FeZn_4(prv)_4(btdd)_3$ after dosing with O_2 at 100 K over the course of gradual warming to 298 K revealed that the material remains highly crystalline under these conditions (fig. S24).

An absorption band was observed to grow in at 831 cm^{-1} upon dosing $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ with O_2 at 100 K, which we assign as an Fe(IV) = O stretch (Fig. 2B, solid lines). This band increased in intensity with heating up to 200 K before diminishing significantly at 250 K and disappearing at 298 K. When the analogous experiment was performed with $^{18}O_2$, the band shifted to 796 cm^{-1}, consistent with a stretching frequency of 794 cm^{-1} calculated for Fe(IV) = ^{18}O using a simple harmonic oscillator model (Fig. 2B, dashed lines). For comparison, the Fe(IV) = ^{16}O and Fe(IV) = ^{18}O stretches in TauD-*J* appear at 821 and 787 cm^{-1}, respectively (*32*). Concomitant with the disappearance of the Fe(IV) = O stretch at 250 K, a new stretch appeared at 3628 cm^{-1} (fig. S17). We attribute this stretch to an Fe(III)–OH species arising from decomposition of the Fe(IV) = O through hydrogen-atom abstraction, possibly from the methyl group of the newly formed acetate ligand (Fig. 2C). At 298 K, a new stretch was apparent at 3678 cm^{-1} (fig. S17), which may correspond to a different coordination environment for the Fe(III)–OH species at higher temperatures. When $^{18}O_2$ was used for dosing, a stretch appeared at 3617 cm^{-1}, in excellent agreement with that calculated for Fe(III)–^{18}OH using a simple harmonic oscillator model (3616 cm^{-1}).

Experimental and computational investigation of the Fe(IV) = O spin state

The species formed upon reaction of $Fe_xZn_{5-x}(prv)_4(btdd)_3$ with dioxygen were further investigated by in situ Mössbauer spectroscopy (see section 2.9.1 of the SM). As discussed above, the 5 K Mössbauer spectrum of desolvated $FeZn_4(prv)_4(btdd)_3$ features a major doublet indicative of high-spin, five-coordinate iron (II) (Fig. 3A and table S6). After dosing with 300 mbar of O_2 at 100 K, this doublet persisted and a new doublet was apparent with δ = 0.260(4) mm/s and $|\Delta E_Q|$ = 0.572(8) mm/s [16.7(2)% area], which we assign to an Fe(IV) = O species. These parameters are similar to those determined for the S = 2 Fe(IV) = O intermediate of TauD (δ = 0.31 mm/s and ΔE_Q = −0.88 mm/s) (*6*, *7*). A detailed comparison with isomer shift and quadrupole splitting values for other reported iron(IV)–oxos reveals that they are more consistent with an S = 2 spin state ($\delta_{S=2}$ = 0.02–0.37 mm/s; $|\Delta E_Q|_{S=2}$ = 0.23–1.27 mm/s) than an S = 1 spin state ($\delta_{S=1}$ < 0.20 mm/s; $|\Delta E_Q|_{S=1}$ = 0.44–2.09 mm/s) (see fig. S40). Density functional theory (DFT) calculations at the B3LYP/def2-TZVP level of theory predict an isomer shift of 0.25 mm/s for an S = 2 six-coordinate Fe(IV) = O moiety featuring acetate coordinated in a κ^2-binding mode, in excellent agreement with the experimental value (tables S6 and S9).

We considered that a slow oxidation rate at 100 K may be a factor limiting the Fe(IV) = O content in O_2-dosed $FeZn_4(prv)_4(btdd)_3$. Indeed, the percent area of the Fe(IV) = O doublet could be increased up to a maximum of 20.0(2)% upon further dosing with of O_2 at 125 and 150 K. (Fig. 3A and table S6). However, subsequent dosing at 163 K did not result in an increase in the percent of the Fe(IV) = O doublet (figs. S19 and

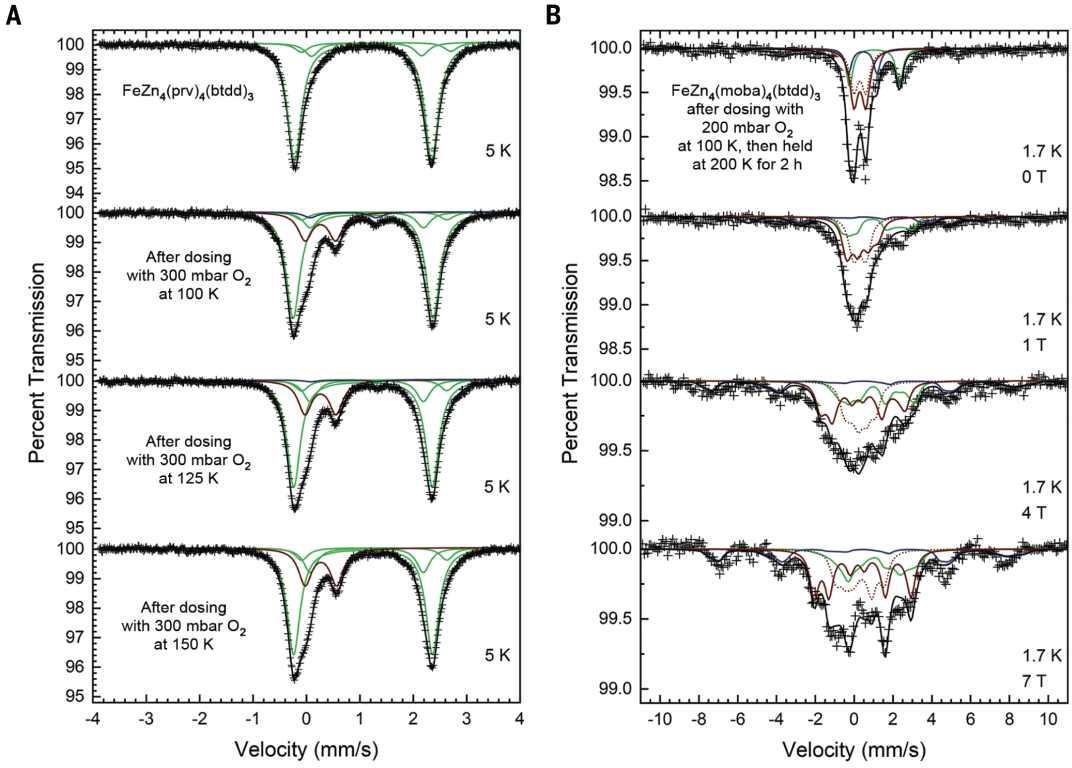

A
FeZn4(prv)4(btdd)3
5 K
After dosing with 300 mbar O2 at 100 K
5 K
After dosing with 300 mbar O2 at 125 K
5 K
After dosing with 300 mbar O2 at 150 K
5 K
Percent Transmission
Velocity (mm/s)
B
FeZn4(moba)4(btdd)3 after dosing with 200 mbar O2 at 100 K, then held at 200 K for 2 h
1.7 K 0 T
1.7 K 1 T
1.7 K 4 T
1.7 K 7 T
Percent Transmission
Velocity (mm/s)

Fig. 3. In situ and variable-field Mössbauer spectra. (**A**) In situ Mössbauer spectra collected at 5 K for desolvated $FeZn_4(prv)_4(btdd)_3$ before and after dosing with 300 mbar of O_2 at 100, 125, and 150 K. All spectra were fit with a minimum number of symmetric quadrupole doublets, all of which have the same linewidth for a given spectrum (see fig. S22 for a timeline of the in situ dosing experiment). Green, blue, and dark red subspectra are consistent with iron(II), iron(III), and iron(IV) species, respectively (**B**) Variable magnetic field Mössbauer spectra collected at 1.7 K and the indicated fields for a sample of $FeZn_4(moba)_4(btdd)_3$ that had been dosed with 100 mbar of O_2 at 100 K, held for 2 hours at 200 K and subsequently dosed with 200 mbar of O_2 at 100 K and then warmed again at 200 K for 2 hours. Green, blue, dark red solid, and dark red dotted subspectra, respectively, were modeled as S = 2 iron(II), S = $^5/_2$ iron(III), isolated S = 2 Fe(IV)=O species, and S = 0 species arising from antiferromagnetic coupling between Fe(IV)=O species within the same node. The spectra were modeled as described in the main text to extract D = 12.7(6) cm^{-1} and A_{iso} = −16.4(4) T for the uncoupled Fe(IV)=O species (see table S13). Note, the width of the spectral splitting for the red trace at 7 T is smaller than has been observed for other S = 2 Fe(IV)=O species in the literature (~5 versus ~8 mm/s, respectively) (*6*, *17*, *41*). However, as demonstrated for another S = 2 Fe(IV)=O complex (*50*), a spectral splitting of about 5 mm/s can be observed if the hyperfine parameters are smaller than the isotropic Fermi-contact contribution of approximately −21 T, as is the case for $FeZn_4(moba)_4(btdd)_3$ (see also fig. S36).

S22). At these temperatures, it is possible that a greater quantity of Fe(IV) = O species is generated initially but that some of these sites react with the acetate ligand to generate a hydroxylated product and reform high-spin iron(II) sites. Analogous intramolecular α-C–H hydroxylation is known for α-ketoglutarate–dependent dioxygenases (*33*), and we discuss this possibility further below. Finally, spectra were collected at 5 K after heating the sample at 250 and 298 K without additional O_2 dosing. After warming at 250 K, the area of the Fe(IV) = O doublet was only 9.7(3)% and after warming at 298 K, the Fe(IV) = O doublet was absent, consistent with the decomposition observed in the DRIFTS data (fig. S20). A Mössbauer spectrum collected at 5 K after ex situ dosing of $FeZn_4(prv)_4(btdd)_3$ with O_2 at 163 K revealed a slightly higher Fe(IV) = O content of 32.2(6)% relative to that achieved under in situ conditions (section 2.9.2 of the SM and fig. S21), which may be due to the different conditions and the nature of the sample used (loose versus compact powder, respectively). The sample remained crystalline after warming to 298 K (fig. S25) and the porosity of the framework did not change significantly (1980 ± 23 m^2/g, fig. S26).

The Mössbauer isomer shift and quadrupole splitting values determined for the Fe(IV) = O species generated in $FeZn_4(prv)_4(btdd)_3$ are more consistent with an S = 2 than an S = 1 spin state, although these parameters alone do not enable an unambiguous assignment. To experimentally assign the spin state, we turned to applied magnetic field Mössbauer spectroscopy (*6*, *17*, *34*). We expected that the low concentration of the Fe(IV) = O species accessible upon dosing $FeZn_4(prv)_4(btdd)_3$ with O_2 would limit data resolution. One plausible explanation for the low detected concentration is decomposition of the Fe(IV) = O species through oxygenation of the newly formed acetate ligand (*33*). Such a side reaction would be precluded if pyruvate were replaced with an α-ketocarboxylate lacking α-C–H bonds. Accordingly, we synthesized $FeZn_4(moba)_4(btdd)_3$, which features a 3,3-dimethyl-2-oxobutyrate ligand with a *tert*-butyl group alpha to the carbonyl (see section 1.8 of the SM). This material is isostructural to $Fe_xZn_{5-x}(prv)_4(btdd)_3$ (fig. S27) and exhibits a comparably high BET surface area (fig. S28). In situ DRIFTS data collected for $FeZn_4(moba)_4(btdd)_3$ upon dosing with O_2 at 100 K support Fe(IV) = O formation through decarboxylation of 3,3-dimethyl-2-oxobutyrate [see fig. S29; ν(Fe = O) = 828 cm^{-1}, ν(Fe = ^{18}O) = 794 cm^{-1}], consistent with data discussed above for $FeZn_4(prv)_4(btdd)_3$.

The zero-field Mössbauer spectrum of $FeZn_4(moba)_4(btdd)_3$ collected at 5 K features a major doublet assigned to high-spin, five-coordinate iron(II) (fig. S31 and table S10). Following a sequence of in situ dosing with O_2 at 100 K and heating at 200 K, a new quadrupole doublet was apparent in the 5 K Mössbauer spectrum with δ = 0.292(1) mm/s and ΔE_Q = −0.603(1) mm/s [61.7(1)% area], assigned to the Fe(IV) = O species (fig. S31; the sign of the quadrupole splitting was determined from variable-field Mössbauer

data, see section 2.12 of the SM). This relative area is significantly larger than the maximum relative area of in situ O_2-dosed $FeZn_4(prv)_4(btdd)_3$ [21.1(1)%], supporting our hypothesis that intramolecular ligand oxygenation may be limiting the Fe(IV) = O content. Variable-field Mössbauer spectra were subsequently collected for O_2-dosed $FeZn_4(moba)_4(btdd)_3$ at 1.7 K under fields ranging from 0 to 7 T and at temperatures from 1.7 to 40 K under a field of 7 T (Fig. 3B and fig. S33; see section 2.12 of the SM). Spectra were also collected for $FeZn_4(moba)_4(btdd)_3$ at temperatures <5 K and fields ranging from 0 to 7 T (fig. S32) to obtain fixed parameters for modeling the residual iron(II) species for the O_2-dosed material (table S12; see section 2.12.2 of the SM).

Consistent with the zero-field Mössbauer spectrum collected at 5 K, the zero-field spectrum for O_2-dosed $FeZn_4(moba)_4(btdd)_3$ collected at 1.7 K could be fit with three subspectra, corresponding to an $S = 2$ iron(II) component, an iron(III) species, and an Fe(IV) = O component (Fig. 3B). The isomer shift of this Fe(IV) = O species is consistent with reported $S = 2$ iron(IV)–oxo species in the literature (fig. S40); however, both $S = 1$ and $S = 2$ models were considered in fitting the variable-field and variable-temperature data. Initial attempts to simultaneously fit the variable-temperature data collected under a 7 T field using this three subspectra model did not fully account for the total spectral intensity. We found that inclusion of a fourth subspectrum, assigned to an $S = 0$ species and likely arising from antiferromagnetically coupled Fe(IV) = O sites within the same cluster node, affords a good fit to the data (see section 2.12.3 of the SM). Although ICP-OES analysis predicts a distribution of one iron site per cluster on average, we cannot exclude a distribution wherein some clusters feature two iron centers per cluster, some feature a single iron center, and other clusters contain zero iron sites.

Simultaneous modeling of the variable-field and variable-temperature spectra was performed through diagonalization of the spin Hamiltonians for the four subspectra to obtain values of the zero-field splitting (D) and hyperfine coupling parameters (A_{xx}, A_{yy}, A_{zz}) for the uncoupled Fe(IV) = O species when assigned as either $S = 2$ or $S = 1$ (see section 2.12.3 of the SM for details and tables S13 to 15). Although the data could be fit with an $S = 2$ or $S = 1$ model (Fig. 3B and figs. S33 to S35), only the results for the $S = 2$ model are chemically reasonable based on a comparison with reported Fe(IV) = O species. The best model for the $S = 2$ Fe(IV) = O species yielded $D = 12.7(6)$ cm^{-1} and $a_{iso} = -16.4(4)$ T (table S13), parameters that are in excellent agreement with those reported for $S = 2$ Fe(IV) = O species ($D_{S=2} = 4$–14 cm^{-1}, $a_{iso(S=2)} = -23$ to -16 T) (*17*, *35*). By contrast, the best model for $S = 1$ Fe(IV) = O sites yielded values that are significantly different from those reported previously for $S = 1$ Fe(IV) = O species (table S14) (*35*–*37*). Furthermore, the signs and relative magnitudes of the spin-dipole contribution $\vec{A}_{SD}$ to the hyperfine coupling tensor are only consistent with an $S = 2$ spin ground state (see section 2.12.4 of the SM).

Additional experimental support for the $S = 2$ assignment for the Fe(IV) = O species was obtained from Fe Kβ x-ray emission spectroscopy (XES; section 2.13 of the SM). Iron Kβ XES involves the measurement of 3p to 1s emissions and because of the Fe 3p–3d exchange contributions to the final state, the relative intensities and energies of the Kβ mainline features—the Kβ′ and $K\beta_{1,3}$ peaks—are diagnostic of the local iron spin state (*38*–*40*). High-spin states typically give rise to more intense Kβ′ features and larger separations between the Kβ′ and $K\beta_{1,3}$ peaks than low-spin states (*39*). The XES spectra for $S = 2$ $FeZn_4(moba)_4(btdd)_3$ and a sample of the framework dosed ex situ with O_2 are given in fig. S39, along with a representative spectrum for a reported molecular compound featuring an $S = 1$ Fe(IV) = O (*40*). The spectra for pristine and O_2-dosed $FeZn_4(moba)_4(btdd)_3$ nearly overlay and feature an intense Kβ′ peak, consistent with the dominant presence of $S = 2$ iron sites in both materials. By contrast, there is no clear Kβ′ peak in the $S = 1$ Fe(IV) = O spectrum, and the $K\beta_{1,3}$ peak appears at a lower energy than in the spectrum for O_2-dosed $FeZn_4(moba)_4(btdd)_3$.

Finally, we used DFT to calculate Mössbauer isomer shifts for the model cluster $Fe(O)(\kappa^2\text{-}OPiv)Zn_4(prv)_3(bta)_6$ ($OPiv^-$, pivalate formed upon $moba^-$ decarboxylation; bta^-, benzotriazolate) featuring an $S = 2$ or $S = 1$ ground state. Geometry optimizations were performed at the BP86-D3BJ/def2-TZVP level of theory. For the $S = 2$ optimized geometry, $\delta_{calc} = 0.27$ mm/s is in excellent agreement with the experimentally determined isomer shift [$\delta = 0.292(1)$ mm/s], whereas the isomer shift calculated for $S = 1$ deviates significantly from the experimental value ($\delta_{calc} = 0.16$ mm/s) (table S11). Furthermore, CASSCF calculations in combination with *N*-electron valence perturbation theory to second order (NEVPT2) support the assignment of an $S = 2$ ground state, which is predicted to be 25 kcal/mol more stable than the lowest energy $S = 1$ configuration (see section 3 of the SM and fig. S41). Altogether, the Mössbauer, Fe Kβ XES, and computational data clearly support the assignment of an $S = 2$ ground state for the Fe(IV) = O sites formed upon oxidation of $FeZn_4(moba)_4(btdd)_3$.

Investigation of the local Fe(IV) = O coordination environment using nuclear resonance vibrational spectroscopy

Nuclear resonance vibrational spectroscopy (NRVS) was used to gain further insight into the local structure of the Fe(IV) = O species (see section 2.7 of the SM). This technique selectively yields the complete set of vibrational modes of Mössbauer-active nuclei and can therefore provide structural insights not accessible using other spectroscopic methods. Figure 4, A and B respectively, show the iron partial vibrational density of states (PVDOS) distributions obtained from data collected at ~100 K for desolvated 95% ^{57}Fe-enriched $FeZn_4(prv)_4(btdd)_3$ before and after ex situ O_2-dosing at 163 K. A new peak at 822 cm^{-1} for the O_2-dosed sample was assigned to an Fe(IV) = O vibration and is similar in magnitude to NRVS peaks reported for other nonheme Fe(IV) = O species in synthetic systems (*41*, *42*, *43*). In support of this assignment, when $^{18}O_2$ was employed for dosing, the vibration appeared instead at 788 cm^{-1} (Fig. 4B, inset; figs. S44 and S47). Consistent results were obtained from PVDOS distributions obtained for ^{57}Fe-enriched $FeZn_4(moba)_4(btdd)_3$ after dosing with O_2 or $^{18}O_2$ (fig. S50), which feature peaks at 820 and 789 cm^{-1}, respectively. The Fe(IV) = O peak was absent in the PVDOS distribution obtained for both frameworks after warming to 298 K (figs. S43 and S51).

Using the truncated cluster $FeZn_4(prv)_4(bta)_6$ as a model for $FeZn_4(prv)_4(btdd)_3$, we performed DFT calculations to simulate the NRVS iron PVDOS for the framework before and after O_2 dosing (Fig. 4, A and B, dark gray curves). The intense stretch at 330 cm^{-1} predicted for the model iron(II) framework corresponds to vibrations associated with bidentate pyruvate binding. Differences in the predicted and experimental intensities likely arise because the cluster model cannot fully describe the phonons of the framework lattice. For the O_2-dosed sample, stretches at 282 and 340 cm^{-1} are assigned as Fe–O vibrations resulting from κ^2-binding of the acetate ligand (Fig. 4C). By contrast, Fe–O vibrations associated with κ^1-binding of acetate are predicted to appear at higher wave numbers (>400 cm^{-1}) (figs. S45 and S46). The calculated Fe(IV) = O stretch is higher than the experimental stretching frequency (919 versus 822 cm^{-1}), likely because of a known systematic overestimation by DFT at these higher energies, which is less pronounced at lower energies (*44*). Finally, simulated NRVS iron PVDOS were also generated for pristine and O_2-dosed $FeZn_4(moba)_4(btdd)_3$ from DFT calculations on the truncated $S = 2$ cluster models $Fe(moba)Zn_4(prv)_3(bta)_6$ and $Fe(O)(\kappa^2\text{-}OPiv)Zn_4(prv)_3(bta)_6$, and the results are in good agreement with the experiment (figs. S48 and S49).

Reactivity studies

We evaluated the reactivity of $Fe_xZn_{5-x}(prv)_4(btdd)_3$ with hydrocarbon substrates in the presence of O_2. Desolvated $FeZn_4(prv)_4(btdd)_3$ was suspended in cyclohexane and exposed to one bar of O_2 at 21°C for 24 hours. Subsequently, CH_3CN-d_3 was added to extract the products, along with CH_2Br_2 as an internal standard. Analysis of the resulting supernatant using ^{1}H

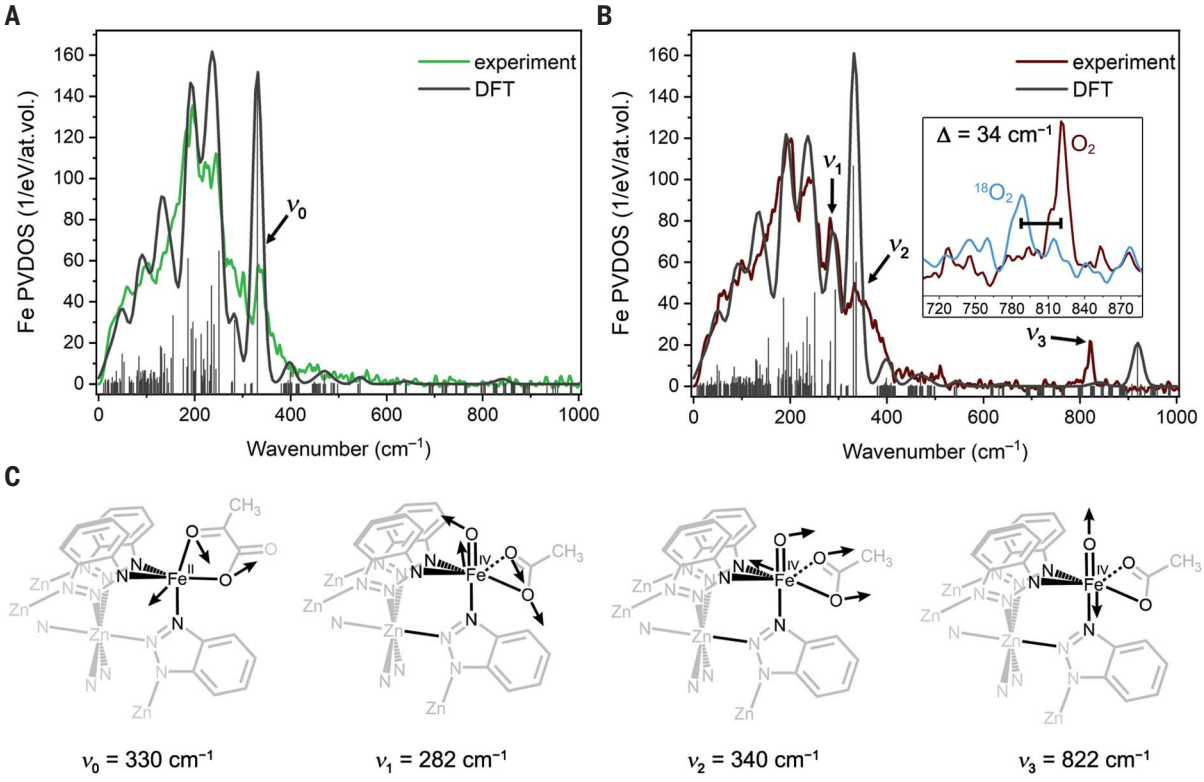


Fig. 4. Nuclear resonance vibrational spectroscopy (NRVS). (**A**) Iron partial vibrational density of states (PVDOS) distribution of ^{57}Fe-enriched, desolvated $FeZn_4(prv)_4(btdd)_3$ from NRVS data collected at ~100 K (green) and DFT computed Fe PVDOS for $FeZn_4(prv)_4(bta)_6$ (dark gray). (**B**) Iron PVDOS distribution of ^{57}Fe-enriched, desolvated $FeZn_4(prv)_4(btdd)_3$ obtained from NRVS data collected at ~100 K after dosing with 200 mbar of O_2 at 163 K (dark red) and DFT-computed PVDOS distribution for ~70% $FeZn_4(prv)_4(bta)_6$ and ~30% of $Fe(O)(\kappa^2\text{-}OAc)Zn_4(prv)_3(bta)_6$ (dark gray). The inset shows a shift of the Fe(IV)=O vibration to lower wave numbers (Δ = 34 cm^{-1}) when $^{18}O_2$ is used (light blue, see also fig. S46). Vertical lines indicate the individual vibrational transitions. (**C**) Assigned vibrational modes of $FeZn_4(prv)_4(btdd)_3$ and the O_2-dosed framework.

NMR spectroscopy and GC-MS revealed the formation of cyclohexanol (22% NMR yield with respect to the iron sites in the framework) with no detectable cyclohexanone (see section 1.9.1 of the SM and figs. S53 and S54). Mössbauer spectroscopy analysis of the framework isolated following this reaction revealed only iron(II) species (fig. S56). The same stoichiometric reaction was also carried out using $FeZn_4(moba)_4(btdd)_3$ in the presence of O_2 (1 bar, 21°C), and analysis of the resulting supernatant using ^{1}H NMR spectroscopy and GC-MS revealed the formation of only cyclohexanol (51% NMR yield) (see section 1.11.1 of the SM and figs. S59 and S60). A stoichiometric control reaction between cyclohexane and $FeZn_4Cl_4(btdd)_3$ in the presence of O_2 did not yield any hydrocarbon oxidation products (figs. S57 and S58).

In order to establish the direct role of the Fe(IV) = O species in C–H oxygenation, we sought to perform a stoichiometric reaction using a framework sample in which the Fe(IV) = O species were generated before the addition of substrate. A sample of $FeZn_4(prv)_4(btdd)_3$ was dosed with 200 mbar O_2 at 163 K and after 2 hours the sample headspace was evacuated and refilled with Ar; a mixture of cyclohexane and CD_2Cl_2 was then added. The suspension was then warmed to 195 K and held for 2 hours and then allowed to warm to 294 K. Under these conditions, no cyclohexane oxidation products were detected through ^{1}H NMR spectroscopy. When a similar reaction was performed using $FeZn_4(moba)_4(btdd)_3$ (see section 1.11.2 of the SM), ^{1}H NMR spectroscopy and GC-MS analysis of the resulting supernatant revealed the formation of cyclohexanone (figs. S61 and S62; 48% NMR yield). The ultimate formation of cyclohexanone in this case—in contrast to cyclohexanol formed in the reaction conducted at 21°C—is attributed to the lower reaction temperature and slower diffusion of cyclohexanol out of the framework pores, which is then further oxidized to cyclohexanone (see section 1.11 of the SM) (*45*).

When the cyclohexane oxidation reaction with $FeZn_4(prv)_4(btdd)_3$ was repeated with the addition of 11 equivalents of pyruvic acid, cyclohexanone and cyclohexanol were obtained in a 2:1 ratio (combined yield of 173%) (see section 1.9.2 of the SM and fig. S64). Powder x-ray diffraction analysis of the solid isolated from this reaction confirmed that the framework remains crystalline (fig. S65). Notably, this result suggests that $FeZn_4(prv)_4(btdd)_3$ can act as a catalyst in hydrocarbon oxidation reactions using the free α-keto acid as a co-substrate, presumably through a similar catalytic cycle as proposed for TauD (fig. S71) (*6*). We found that acetic acid byproduct is formed in this reaction in 288% yield with respect to the iron sites, which suggests 115% of pyruvic acid conversion to unidentified products. Such unproductive turnover at some of the iron sites is consistent with the relatively low yield in the stoichiometric reaction. A number of studies of synthetic and enzymatic systems have established that C–H bond activation by Fe(IV) = O species proceeds through H-atom abstraction, as evidenced by large primary kinetic isotope effects (KIEs) (*45*, *46*). Consistent with these results, we determined an intermolecular competition KIE value of 29.8 ± 1.0 from the reaction of $FeZn_4(prv)_4(btdd)_3$ with a mixture of cyclohexane and cyclohexane-d_{12} under an atmosphere of O_2 (see section 1.10 of the SM and figs. S66 and S67).

The reactivity of $Fe_{1.8}Zn_{3.2}(prv)_4(btdd)_3$ in the presence of O_2 was also evaluated using gaseous ethane as a substrate in a high-pressure batch reactor (section 1.12 of the SM). Using a high-pressure mixture of ethane and dioxygen, ethanol and acetaldehyde were obtained in a 3:1 ratio with a high combined yield of 82% (fig. S68). This yield is much higher than that obtained in

the stoichiometric cyclohexane oxidation, likely due to the high ethane concentration close to the iron sites under high pressure. Although there are enzymatic systems capable of oxidizing ethane to ethanol with O_2 (*47*), our result is the first synthetic example of ethane oxidation through an unambiguously characterized $S = 2$ Fe(IV) = O intermediate generated by O_2.

Outlook

We have developed the frameworks $Fe_xZn_{5-x}(prv)_4(btdd)_3$ and $FeZn_4(moba)_4(btdd)_3$ featuring iron(II) sites that activate O_2 at 100 K to form high-spin Fe(IV) = O species, reactivity that is reminiscent of O_2 activation in TauD. These frameworks are rare non-enzymatic systems capable of catalytic hydrocarbon oxygenation, including the oxidation of ethane to ethanol, through a reactive high-spin Fe(IV) = O intermediate generated from dioxygen. This work establishes a foundation for the development of iron-containing metal–organic frameworks as heterogeneous catalysts that closely mimic the reactivity of metalloenzymes.

REFERENCES AND NOTES

1. B. G. Hashiguchi *et al*., *Science* **343**, 1232–1237 (2014).
2. J. Shan, M. Li, L. F. Allard, S. Lee, M. Flytzani-Stephanopoulos, *Nature* **551**, 605–608 (2017).
3. V. L. Sushkevich, D. Palagin, M. Ranocchiari, J. A. van Bokhoven, *Science* **356**, 523–527 (2017).
4. E. G. Kovaleva, J. D. Lipscomb, *Nat. Chem. Biol.* **4**, 186–193 (2008).
5. C. Krebs, D. Galonić Fujimori, C. T. Walsh, J. M. Bollinger Jr., *Acc. Chem. Res.* **40**, 484–492 (2007).
6. J. C. Price, E. W. Barr, B. Tirupati, J. M. Bollinger Jr., C. Krebs, *Biochemistry* **42**, 7497–7508 (2003).
7. S. Sinnecker *et al*., *J. Am. Chem. Soc.* **129**, 6168–6179 (2007).
8. J. Hohenberger, K. Ray, K. Meyer, *Nat. Commun.* **3**, 720 (2012).
9. M. P. Mehn, K. Fujisawa, E. L. Hegg, L. Que Jr., *J. Am. Chem. Soc.* **125**, 7828–7842 (2003).
10. A. Thibon *et al*., *Angew. Chem. Int. Ed.* **47**, 7064–7067 (2008).
11. F. Li, K. M. Van Heuvelen, K. K. Meier, E. Münck, L. Que Jr., *J. Am. Chem. Soc.* **135**, 10198–10201 (2013).
12. D. Kass *et al*., *J. Am. Chem. Soc.* **142**, 5924–5928 (2020).
13. M. A. Dedushko, M. B. Greiner, A. N. Downing, M. Coggins, J. A. Kovacs, *J. Am. Chem. Soc.* **144**, 8515–8528 (2022).
14. J. P. Bigi *et al*., *J. Am. Chem. Soc.* **134**, 1536–1542 (2012).
15. M. Puri, L. Que Jr., *Acc. Chem. Res.* **48**, 2443–2452 (2015).
16. C. E. MacBeth *et al*., *Science* **289**, 938–941 (2000).
17. J. B. Gordon *et al*., *J. Am. Chem. Soc.* **143**, 21637–21647 (2021).
18. B. E. R. Snyder *et al*., *Nature* **536**, 317–321 (2016).
19. E. Tabor *et al*., *Sci. Adv.* **6**, eaaz9776 (2020).
20. B. Rungtaweevoranit *et al*., *ACS Appl. Mater. Interfaces* **15**, 26700–26709 (2023).
21. A. M. Wright *et al*., *Chem* **4**, 2894–2901 (2018).
22. C. E. Bien *et al*., *J. Am. Chem. Soc.* **140**, 12662–12666 (2018).
23. J. G. Vitillo, C. C. Lu, A. Bhan, L. Gagliardi, *Cell Rep. Phys. Sci.* **4**, 101422 (2023).
24. H.-C. Zhou, J. R. Long, O. M. Yaghi, *Chem. Rev.* **112**, 673–674 (2012).
25. H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **341**, 1230444 (2013).
26. D. J. Xiao *et al*., *Nat. Chem.* **6**, 590–595 (2014).
27. A. Iliescu, J. J. Oppenheim, C. Sun, M. Dincă, *Chem. Rev.* **123**, 6197–6232 (2023).
28. A. Shaabani, R. Mohammadian, H. Farhid, M. K. Alavijeh, M. M. Amini, *Catal. Lett.* **149**, 1237–1249 (2019).
29. D. Denysenko, J. Jelic, K. Reuter, D. Volkmer, *Chemistry* **21**, 8188–8199 (2015).
30. D. Denysenko *et al*., *Chemistry* **17**, 1837–1848 (2011).
31. P. Gütlich, E. Bill, A. X. Trautwein, *Mössbauer Spectroscopy and Transition Metal Chemistry* (Springer, 2011).
32. D. A. Proshlyakov, T. F. Henshaw, G. R. Monterosso, M. J. Ryle, R. P. Hausinger, *J. Am. Chem. Soc.* **126**, 1022–1023 (2004).
33. V. F. Oswald *et al*., *J. Am. Chem. Soc.* **142**, 11804–11817 (2020).
34. K. Warm *et al*., *Angew. Chem. Int. Ed.* **60**, 6752–6756 (2021).
35. M. G. Thompson *et al*., *Nat. Commun.* **11**, 2931 (2020).
36. V. Ksenofontov *et al*., *Hyperfine Interact.* **141**, 47–52 (2002).
37. A. R. McDonald, L. Que Jr., *Coord. Chem. Rev.* **257**, 414–428 (2013).
38. T. A. Jackson *et al*., *J. Am. Chem. Soc.* **130**, 12394–12407 (2008).
39. J. England *et al*., *Angew. Chem. Int. Ed.* **48**, 3622–3626 (2009).
40. N. Lee, T. Petrenko, U. Bergmann, F. Neese, S. DeBeer, *J. Am. Chem. Soc.* **132**, 9715–9727 (2010).
41. C. J. Pollock, M. U. Delgado-Jaime, M. Atanasov, F. Neese, S. DeBeer, *J. Am. Chem. Soc.* **136**, 9453–9463 (2014).
42. G. E. Cutsail 3rd *et al*., *Eur. J. Biochem.* **27**, 573–582 (2022).
43. C. B. Bell 3rd *et al*., *Angew. Chem. Int. Ed.* **47**, 9071–9074 (2008).
44. A. C. Weitz *et al*., *Angew. Chem. Int. Ed.* **57**, 16010–16014 (2018).
45. B. E. R. Snyder *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4565–4570 (2018).
46. A. N. Biswas *et al*., *J. Am. Chem. Soc.* **137**, 2428–2431 (2015).
47. J. C. Price, E. W. Barr, T. E. Glass, C. Krebs, J. M. Bollinger Jr., *J. Am. Chem. Soc.* **125**, 13008–13009 (2003).
48. P. Meinhold, M. W. Peters, M. M. Y. Chen, K. Takahashi, F. H. Arnold, *ChemBioChem* **6**, 1765–1768 (2005).
49. J. Börgel *et al*., "Powder x-ray diffraction and iron Kβ x-ray emission spectroscopy data," Dryad (2023); https://doi.org/10.5061/dryad.wh70rxwtp.

ACKNOWLEDGMENTS

We thank K. M. Carsch, T. D. Harris, A. Nguyen, J. Martell, R. Murphy, J. L. Peltier, M. Qi, B. E. R. Snyder, D. X. Ngo, X. Shen, and Y. Yang for helpful discussions. **Funding:** This research was supported by the US Department of Energy Office of Basic Energy Sciences under award DE-SC0019992. Single-crystal x-ray diffraction data were collected at beamlines 11.3.1 and 12.2.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory, a user facility supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences & Biosciences Division Heavy Element Chemistry Program under contract DE-AC02-05CH11231. Synchrotron powder x-ray diffraction data were collected on the 17-BM-B Beamline at the Advanced Photon Source, a US Department of Energy Office of Science User Facility operated by Argonne National Laboratory. NRVS data was collected at beamline 3-ID-B at the Advanced Photon Source. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences (contract DE-AC02-06CH11357). Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, US Department of Energy (contract DE-AC02-05CH11231). Instruments in the UC Berkeley College of Chemistry NMR facility are supported in part by NIH S10OD024998. EPR spectroscopic studies were funded by the National Institutes of Health (NIH 1R35GM126961 to R.D.B.). R.D.B. acknowledges support from the Miller Institute of Basic Research in Science, University of California, Berkeley. The computing cluster at the Molecular Graphics and Computation Facility at UC Berkeley is supported by NIH S10OD023532. The Fe Kβ XES measurements were carried out at the PINK beam line at the BESSY II electron storage ring operated by the Helmholtz-Zentrum Berlin für Materialien und Energie. D.J.S., S.P., E.B., D.A.P., S.D., and F.N. acknowledge the Max-Planck-Gesellschaft for funding. J.B. acknowledges the Deutsche Forschungsgemeinschaft for a Postdoctoral Research Fellowship. R.C.R. acknowledges the NASA Space Technology Graduate Research Opportunities fellowship. J.O. acknowledges the National Institute of General Medical Sciences for a Postdoctoral Research Fellowship under award F32GM143925. V.C. and M.V. acknowledge the Italian Ministry of University and Research for a Ph.D. fellowship awarded to M.V. and funding through PRIN2017 program (Project "Moscato" n° 2017KKP5ZR_004). **Author contributions:** K.H., J.B., and J.R.L. formulated the project. K.H. and J.B. synthesized the materials, collected and analyzed reactivity study data, and collected and analyzed the gas adsorption data. H.K., J.W.T., and A.B.T. collected and analyzed the single-crystal x-ray diffraction data. H.Z., H.M., X.L., and J.A.R. collected and analyzed the solid-state NMR spectra. D.J.S. collected the applied-field Mössbauer data. D.J.S., K.H., J.B., K.R.M., F.G., G.J.L., and F.N. analyzed the data. D.J.S. and E.B. collected and analyzed the MCD data. K.C. and R.D.B. collected and analyzed EPR data. C.D. and J.J.U. collected and analyzed the EDX data. Z.Z. collected and analyzed the thermogravimetric analysis data. J.G.P. collected and analyzed the magnetic susceptibility data. H.Z.H.J., K.H., and J.B. collected and analyzed the IR spectra. M.V.P., H.Z.H.J., K.H., J.B., J.O., M.V., and V.C. collected and analyzed the PXRD data. J.B., K.H., E.E.A., J.Z., M.Y.H., and B.L. collected and analyzed the NRVS data. K.H., J.B., R.C.R., F.G., and G.J.L. collected and analyzed the Mössbauer spectroscopy data. J.B., K.H., D.A.P., and F.N. performed the electronic structure calculations and analyzed the results. S.P. and S.D. collected and analyzed the Fe Kβ XES data. K.H., J.B., K.R.M., and J.R.L. wrote the manuscript, and all authors contributed to revising the manuscript. **Competing interests:** The University of California, Berkeley, has applied for a patent (International application no. PCT/US23/70362) on the synthesis of the iron containing metal–organic frameworks and their reactivity with hydrocarbon substrates in the presence of dioxygen, on which J.B., K.H., and J.R.L. are listed as coinventors. **Data and materials availability:** The supplementary materials contain complete experimental and spectral details for all new compounds reported herein. Single-crystal x-ray diffraction data have been made available free of charge from the Cambridge Crystallographic Data Centre under reference numbers CCDC 2166411 and 2166412. Powder x-ray diffraction and iron Kβ x-ray emission spectroscopy data are deposited in Dryad (*49*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add7417
Materials and Methods
Supplementary Text
Figs. S1 to S71
Tables S1 to S17
References (*50*–*106*)
Data S1 and S2

Submitted 30 June 2022; resubmitted 15 June 2023
Accepted 24 September 2023
10.1126/science.add7417

ACTIVE GALAXIES

Supermassive black hole feeding and feedback observed on subparsec scales

Takuma Izumi[1,2,3]*, Keiichi Wada[4,5,6], Masatoshi Imanishi[1,3], Kouichiro Nakanishi[1,3], Kotaro Kohno[7,8], Yuki Kudoh[1,4,9], Taiki Kawamuro[10], Shunsuke Baba[4], Naoki Matsumoto[9], Yutaka Fujita[2], Konrad R. W. Tristram[11]

Active galaxies contain a supermassive black hole at their center that grows by accreting matter from the surrounding galaxy. The accretion process in about the central 10 parsecs has not been directly resolved in previous observations because of the small apparent angular sizes involved. We observed the active nucleus of the Circinus Galaxy using submillimeter interferometry. A dense inflow of molecular gas was evident on subparsec scales. We calculated that less than 3% of this inflow is accreted by the black hole, with the rest being ejected by multiphase outflows, providing feedback to the host galaxy. Our observations also reveal a dense gas disk surrounding the inflow that is gravitationally unstable, which drives the accretion into about the central 1 parsec.

Supermassive black holes (SMBHs), those with masses $M_{BH} > 10^6\ M_\odot$ (where $M_\odot$ is the mass of the Sun), are ubiquitous in the centers of nearby galaxies (*1*, *2*). SMBHs grow by accreting mass from the surrounding galaxy (*3*), which releases energy that can be observed as an active galactic nucleus (AGN) (*4*). Major mergers of gas-rich galaxies trigger intense SMBH accretion and bursts of star formation (*5*, *6*). More gradual (secular) processes such as perturbations caused by nonaxisymmetric gravitational potentials and minor mergers (*7*, *8*) are predicted to be sufficient to sustain lower-luminosity AGNs, such as those classified as Seyfert galaxies.

Previous observations have shown that material can be efficiently supplied from the interstellar medium of an entire galaxy to its central region [the central ~100 parsecs (pc), ~1% of the size of the galaxy] (*9*). Resolved observations of this central region using molecular and atomic lines (*10*–*12*) have shown that the accumulated gas often forms a circumnuclear disk (CND) because of its retained angular momentum. Submillimeter observations have been used to study the parsec-scale molecular gas in nearby galaxies (*13*, *14*). However, little is known about the supply of material to the innermost tens of parsecs [particularly within a radius from the center (r) < 10 pc] in AGNs, where SMBHs usually dominate the gravitational potential (*2*). An exception is the nearby active galaxy NGC 1068, in which the subparsec-scale H_2O maser disk and the ~10-pc scale molecular torus have been found to counter-rotate, and this was proposed to enhance mass accretion onto its SMBH (*15*, *16*). However, similar dynamics have not been observed in other AGNs, so they might not be common (*11*). Previous observations of cold gas have not provided quantitative estimates of the mass accretion rate at r < 10 pc.

Absorption line measurements provide complementary information on the supply of gas to SMBHs because they are sensitive to gas clouds along our line of sight toward the bright AGN. These can directly determine inflow and outflow velocities from the red and blue shift of each line relative to the systemic velocity. Absorption line observations have shown that some AGNs have inflows of gas (*17*, *18*). However, whereas some absorbers have been interpreted as being located near the central AGN (r < 100 pc), this interpretation has been challenged (*19*) because those studies were limited by the spatial resolution (~100 to 1000 pc) of the observations.

AGN-driven feedback affects star formation in the host galaxy (*20*, *21*). The supply of material to the SMBH is known to trigger this feedback, although it has not been possible to quantitatively connect the gas supply to the subsequent feedback, especially the multiphase structure and dynamics of the gas at the parsec scale around AGN (*9*, *20*).

Observations of the Circinus Galaxy

To investigate parsec-scale feeding and feedback, we considered the Circinus Galaxy (also known as ESO 97-13), a nearby AGN (distance 4.2 ± 0.7 Mpc) (*22*). The active nucleus of this galaxy has a moderate intrinsic x luminosity (at energies 2 to 10 keV) of (5) × 10^{42} erg s^{-1}, which is heavily obscured by gas and dust (*23*). Around the SMBH, there is a subparsec-scale gas disk, probed by H_2O maser emission, that has been used to measure the dynamical mass of the SMBH as (1.7 ± 0.3) × $10^6\ M_\odot$ (*24*). Other observations have shown a one-sided outflowing cone of ionized gas that extends for up to a kiloparsec (*25*–*27*). Previous submillimeter observations of carbon monoxide (CO) and atomic carbon ([C I]) showed a gas-rich CND (r ~ 30 pc) with a denser gas concentration toward the inner 10 pc (*11*, *28*). High-resolution ionized gas observations have been performed in the rest-frame optical and near-infrared wavelengths (*25*–*27*), but they are affected by dust extinction in the nucleus region.

We performed observations of the Circinus Galaxy with the Atacama Large Millimeter/submillimeter Array (ALMA) radio telescope (*29*) at a spatial resolution of 0.5 to 2.6 pc. We supplemented our observations with additional archival ALMA data. The combined dataset targeted emission lines of hydrogen cyanide HCN (J = 3→2, where J is the rotational quantum number) at rest frequency (ν_{rest}) = 265.8864 GHz, a recombination line of atomic hydrogen (H36α) at ν_{rest} = 135.2860 GHz, CO (J = 3→2) at ν_{rest} = 345.7960 GHz, and [C I] $^3P_1$–$^3P_0$ at ν_{rest} = 492.1607 GHz. For each line, we calculated the critical density (n_{cr}) at which the molecule can be excited to the upper energy state of the line by collisions with H_2 molecules, in the optically thin limit as $n_{cr} = A_{ul}/C_{ul}$, where A_{ul} and C_{ul} are the Einstein A and C coefficients of a transition between the upper state (u) and lower state (l). HCN (J = 3→2) traces very dense molecular gas (n_{cr} = 5.8 × 10^6 cm^{-3} at 50 K), so we expected it to probe parsec-scale dense flows more effectively than low-J CO lines (which have n_{cr} ~ 10^3 to 10^4 cm^{-3}). We expected the H36α line to trace feedback flows of ionized gas. The CO (J = 3→2) and [C I] $^3P_1$–$^3P_0$ lines traced medium-density molecular gas (n_{cr} = 1.1 × 10^4 cm^{-3} at 50 K) and diffuse atomic gas (n_{cr} = 3.7 × 10^2 cm^{-3}), respectively. These submillimeter lines allowed us to study the parsec-scale gas in multiple phases and densities without it being obscured by dust extinction.

Figure 1 shows the [C I] $^3P_1$–$^3P_0$ and CO (J = 3→2) observations, which trace the global spiral arms of the host galaxy that converge to form the CND. Compared with CO (J = 3→2), [C I] $^3P_1$–$^3P_0$ is more concentrated in the central region (Fig. 1A), perhaps caused by dissociation of CO into C by the x-ray radiation from the AGN (a central x-ray–dominated region) (*29*–*31*). The map shows a parsec-scale compact concentration of HCN (J = 3→2) in the inner region (Fig. 1C; the central hole is caused by continuum absorption). We expected such dense

[1]National Astronomical Observatory of Japan, Tokyo 181-8588, Japan. [2]Department of Physics, Graduate School of Science, Tokyo Metropolitan University, Tokyo 192-0397, Japan. [3]Department of Astronomical Science, The Graduate University for Advanced Studies, SOKENDAI, Tokyo 181-8588, Japan. [4]Graduate School of Science and Engineering, Kagoshima University, Kagoshima 890-0065, Japan. [5]Research Center for Space and Cosmic Evolution, Ehime University, Ehime 790-8577, Japan. [6]Faculty of Science, Hokkaido University, Hokkaido 060-0810, Japan. [7]Institute of Astronomy, Graduate School of Science, The University of Tokyo, Tokyo 181-0015, Japan. [8]Research Center for the Early Universe, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan. [9]Astronomical Institute, Tohoku University, Miyagi 980-8578, Japan. [10]Cluster for Pioneering Research, RIKEN, Saitama 351-0198, Japan. [11]European Southern Observatory, Vitacura, Santiago 19001, Chile.
*Corresponding author. Email: takuma.izumi@nao.ac.jp

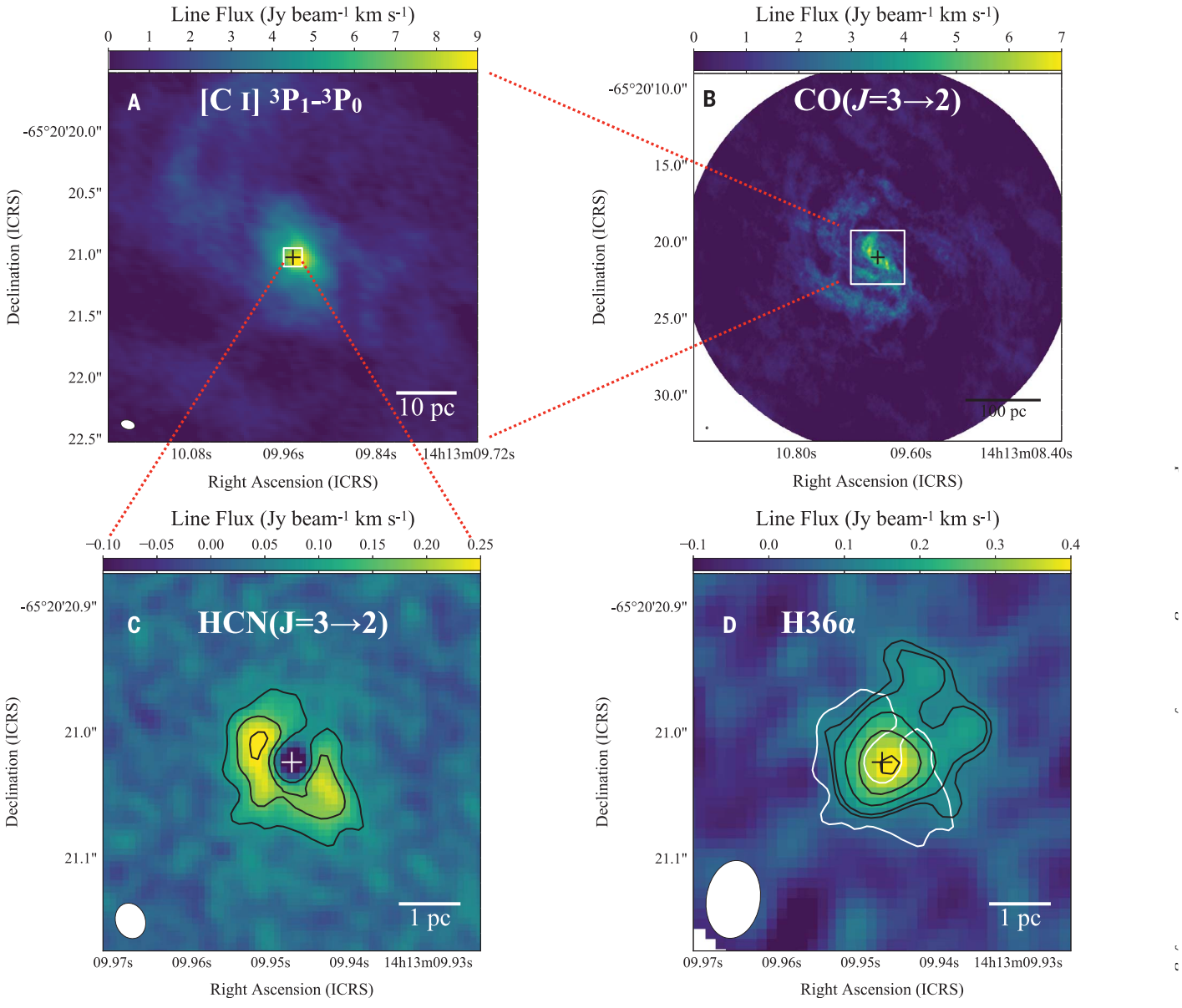


Fig. 1. Observed spatial distributions of multiple gas phases. (**A**) Atomic gas traced by [C I] $^3P_1$–$^3P_0$ (uncertainty 1σ = 0.088 Jy beam^{-1} km s^{-1}) in the central 7 arcsec (~140 pc) of the Circinus Galaxy. The white box indicates the region shown in (C) and (D). (**B**) Medium-density gas traced by CO (J = 3→2) (1σ = 0.055 Jy beam^{-1} km s^{-1}) on larger scales. The white square indicates the region shown in (A). (**C**) Dense molecular gas traced by HCN (J = 3→2) (1σ = 0.017 Jy beam^{-1} km s^{-1}; contours are drawn at 5σ, 10σ, and 15σ). (**D**) Ionized gas traced by H36α (black contours, drawn at 3σ, 3.5σ, 5σ, 7σ, and 10σ, where 1σ = 0.040 Jy beam^{-1} km s^{-1}). The white contour is the 5σ contour of HCN (J = 3→2) from (C). In all panels, the plus sign indicates the AGN position and the ellipses at the bottom left indicate the synthesized beam sizes. The coordinates are expressed in the International Celestial Reference System (ICRS).

molecular gas to survive the x-ray irradiation because it is predominantly in the shielded midplane of the disk (the proposed geometrical structure is shown in Fig. 2E). Lower-density gas (traced by low-J CO) is located above the midplane, so it is more easily dissociated. The parsec-scale distribution of H36α traces the ionized gas without severe dust extinction (Fig. 1D). A one-sided hornlike structure extends almost perpendicular to the HCN disk, toward the northwest (the middle position angle of the emission extended to the northwest is 313°, measured counterclockwise from north). This direction, as well as the half-opening angle (~33° to 39°) of the H36α cone, are consistent with the larger kiloparsec-scale ionization cone seen in previous observations (*25*, *26*).

The parsec-scale inflow rate

To estimate the mass inflow rate onto the SMBH, we first needed to constrain the geometry of the disk. Figure 2A shows that the velocity pattern of the CND is dominated by rotation. We used this map to measure the position angles of the CND-scale kinematic major and minor axes, finding 210° and 300°, respectively. The parsec-scale velocity pattern in the HCN (J = 3→2) map (Fig. 2B) indicates rotation following Keplerian orbits. This motion is evident within 0.1 arcsec (equivalent to 2 pc) from the SMBH in position-velocity diagrams (PVDs) of both HCN (J = 3→2) and [C I] $^3P_1$–$^3P_0$ measured along the CND-scale major axis (Fig. 2C). This Keplerian motion is more prominent in the blue approaching side of the HCN disk than the red receding side, implying asymmetric gas mass distribution at the parsec scale.

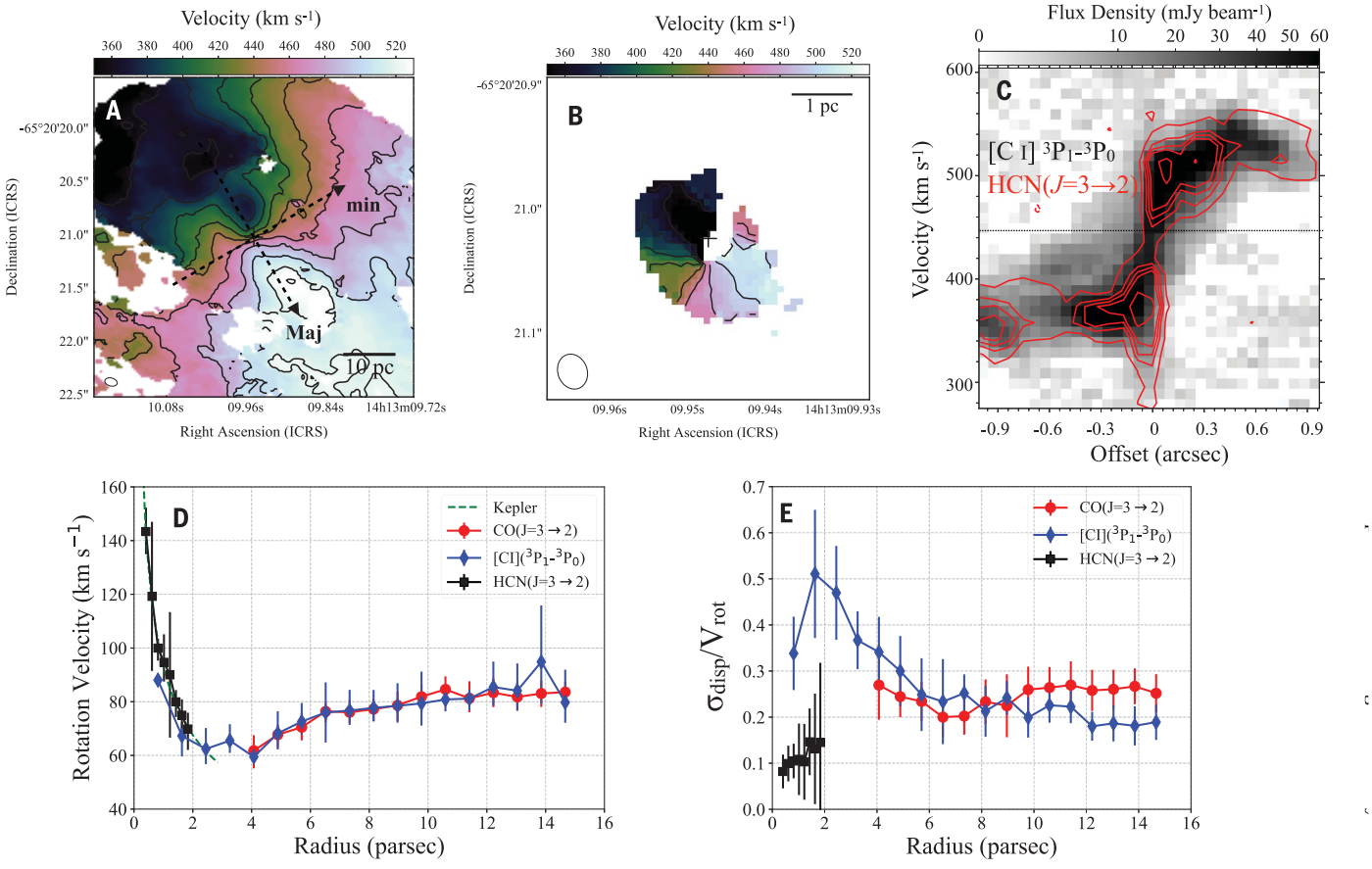


Fig. 2. Kinematic properties of the observed gas. (**A**) Line-of-sight velocity (color bar) measured from the [C I] $^3P_1$–$^3P_0$ data, with dashed arrows indicating the kinematic major and minor axes as labeled. Fig. S3A shows an equivalent plot for CO (J = 3→2). (**B**) Parsec-scale velocity field measured from the HCN (J = 3→2) data. The absorption component close to the AGN was masked out. In (A) and (B), black contours start at 360 km s^{-1} and are separated by steps of 20 km s^{-1}. Black ellipses at the lower left indicate the beam sizes. (**C**) Major-axis PVDs of [C I] $^3P_1$–$^3P_0$ (grayscale, 1σ = 1.26 mJy beam^{-1}) and beam-matched HCN (J = 3→2) (red contours at −3σ, 3σ, 5σ, 6σ, 7σ, and 10σ, where 1σ = 0.68 mJy beam^{-1}). The horizontal dashed line indicates the systemic velocity of 446 km s^{-1}. (**D**) Rotation curves of each of the gas tracers, as indicated in the legend. The green dashed line is a model of Keplerian motion due to a central black hole of mass 2.0 × 10^6 $M_\odot$, which was fitted to the data at r < 2 pc. (**E**) Radial profile of the σ_{disp}/V_{rot} ratio, which is a proxy for the disk aspect ratio, for the same gas tracers. Near the AGN, the HCN (J = 3→2) disk is much thinner than the [C I] $^3P_1$–$^3P_0$ disk. In panels (D) and (E), error bars indicate 1σ uncertainty.

We modeled the three-dimensional structure of these multiphase gases using a tilted-ring scheme (*32*) that decomposes the rotation velocity (V_{rot}) and random dispersion (σ_{disp}) (*29*). Figure 2D shows the decomposed rotation curves of CO (J = 3→2), [C I] $^3P_1$–$^3P_0$, and HCN (J = 3→2) in the central r < 15 pc. The rotation curves of CO (J = 3→2) and [C I] $^3P_1$–$^3P_0$ are consistent with each other and smoothly connect to those previously measured over larger scales of r = 10 to 150 pc (*11*). The [C I] $^3P_1$–$^3P_0$ and HCN (J = 3→2) rotation curves indicate a Keplerian rotation pattern following a power law (velocity ∝ r^n) with index n = 0.46 ± 0.06. We estimated the SMBH mass from these curves, finding (2.0 ± 0.1) × 10^6 $M_\odot$, which is consistent with previous H_2O maser observations (*24*).

Figure 2E shows the radial profiles of the ratio σ_{disp}/V_{rot} for each line, which are proxies for the aspect ratio (scale height to radius ratio) of the disk, assuming hydrostatic equilibrium. In the innermost region, within a few parsecs of the SMBH, the diffuse atomic gas (traced by [C I]) forms a geometrically thick structure, whereas the dense molecular gas (traced by HCN) is confined in a thin disk. This confirms our expectation that the dense molecular gas is predominantly in the midplane of the disk, where inflows toward the SMBH are expected (*33*).

We identified inflow motion in the HCN (J = 3→2) spectrum measured at the position of the AGN (Fig. 3A), which displays a deep continuum absorption feature close to the systemic velocity (V_{sys}), as well as emission at velocities < 390 km s^{-1}. This absorption feature is skewed toward the redder side of V_{sys}, consistent with an inverse P-Cygni profile (a redder absorption and bluer emission pattern), which is characteristic of inflows. Because the dense gas forms a thin disk, and this disk is nearly edge-on (at least at the scales of the H_2O maser disk), we fitted this profile with a two-layer infall model (*29*, *34*), which indicated an inflow velocity of 7.4 ± 1.0 km s^{-1}. Assuming that the volume-averaged gas density n_{H2} equals the range of HCN (J = 3→2) critical densities [n_{cr} = (5.8 to 10) × 10^6 cm^{-3} over a temperature range of 20 to 50 K], we estimated the mass inflow rate as ~0.20 to 0.34 $M_\odot$ year^{-1} at r = 0.27 pc (half the spatial resolution). This inflow rate is sufficient to sustain the luminosity of the AGN [the SMBH accretion rate is ~0.006 $M_\odot$ year^{-1} (*29*)].

An alternative interpretation of the HCN (J = 3→2) absorption is that it is caused by a dense cloud on an elliptical orbit around the SMBH. Such elliptical orbits have been frequently observed on larger galactic scales but become less stable in the inner part of a galaxy because of the many overlapping gas cloud orbits. If the system is globally stable, then frequent collisions and tidal interactions between gas clouds would disrupt the clouds. As a result, we expected a smooth circumnuclear disk with

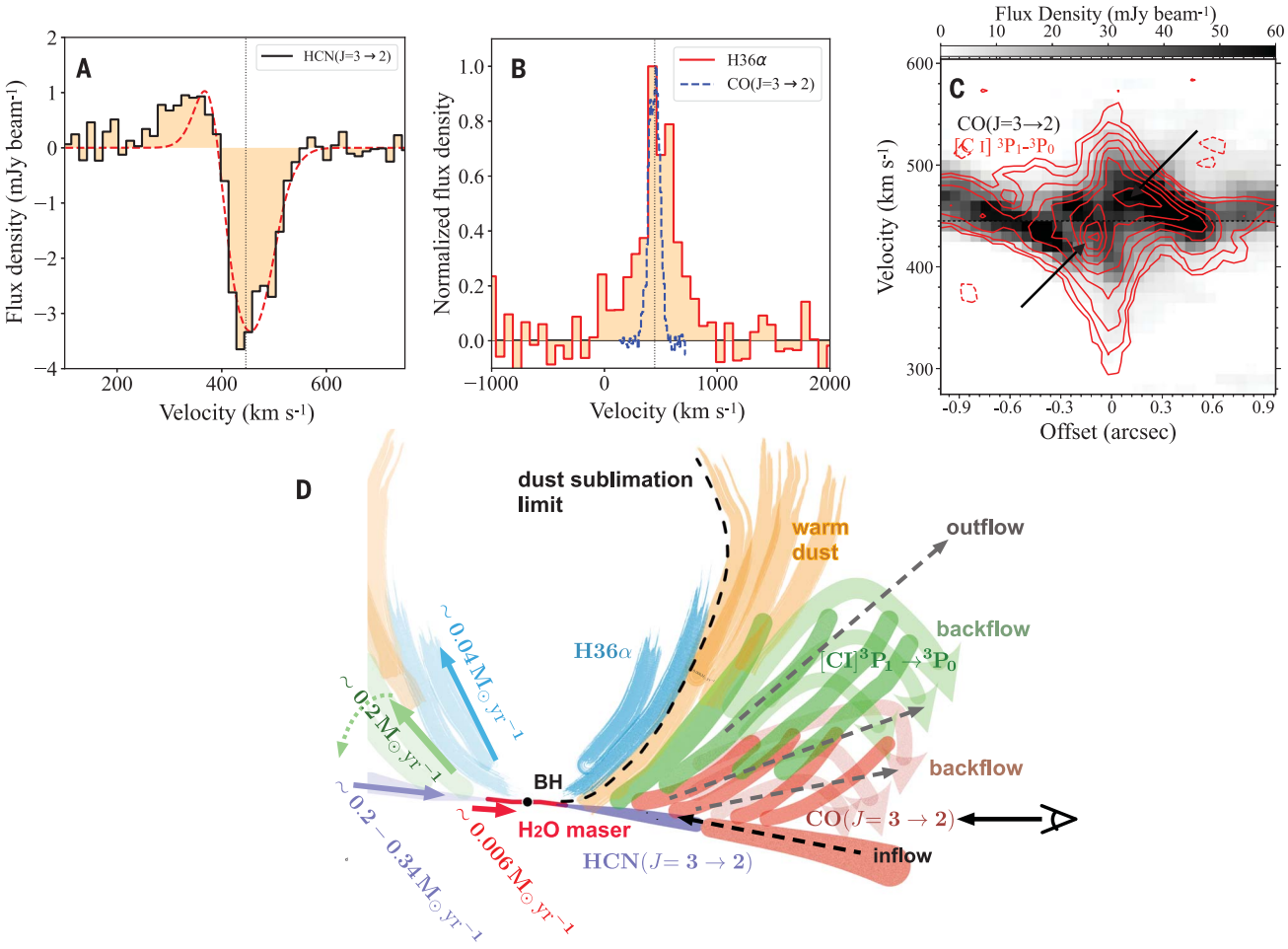


Fig. 3. Spectra at the AGN position and our inferred geometry of the multiphase gas flows. (**A**) HCN (J = 3→2) spectrum (black histogram with orange shading) measured at the AGN position, overlain with an infall model (dashed red line) (*29*). (**B**) H36α (red histogram with orange shading) and CO (J = 3→2) (blue dashed line) spectra at the same position, normalized by their peak fluxes. The Gaussian FWHM is 393 ± 44 km s^{-1} for H36α and 115 ± 1 km s^{-1} for CO (J = 3→2). In panels A and B, vertical dotted lines indicate V_{sys}. (**C**) Minor-axis PVDs of CO (J = 3→2) (grayscale, 1σ = 0.78 mJy beam^{-1}) and [C I] $^3P_1$–$^3P_0$ (red contours drawn at −3σ, 3σ, 5σ, 10σ, 20σ,..., 90σ, where 1σ = 1.26 mJy beam^{-1}). The black arrows indicate outflow features discussed in the text. (**D**) Schematic diagram showing a cross-section of the geometry we infer for the parsec-scale multiphase gas flows. The large scale spiral arms converge to form the CND we observe in CO (J = 3→2) (thick red bands). At the central parsec scale, there is a dense molecular gas disk (purple bands) traced by HCN (J = 3→2), which shows inflow motion (black dashed arrow). This dense disk connects to the H_2O maser disk at the central subparsec scale (thin red line). Fast ionized out flows observed in H36α (blue bands) and slow atomic outflows observed in [C I] $^3P_1$–$^3P_0$ (green bands) occur above the disk. The atomic outflow is slower so will fall back to the disk (labeled "backflow"). Previously reported mid-infrared dust elongation toward the polar direction (*33*) is interpreted as dust entrained by these winds (orange bands). The dense molecular disk is gravitationally unstable in the inner parsec (Fig. 4), which drives the accretion and consequent feedback. Estimated mass flow rates in each phase (colored arrows) are labeled on the left of the panel. The black arrow indicates our line of sight.

density fluctuations, with nearly circular rotation, and with self-regulated turbulence contributing to the viscosity of the gas (*35*). In the central parsec-scale region of the Circinus Galaxy, the expected mean free path of the clouds is much smaller than the disk size, so any clouds on elliptical orbits would collide with another cloud or be tidally disrupted many times within even a single orbit (*29*). The HCN (J = 3→2) rotation curve is consistent with Keplerian rotation, implying that most orbits in this region are circular. We saw no evidence of a nuclear bar, which could introduce elliptical motions. Therefore, we considered the elliptical orbit scenario to be unlikely and thus concluded that the HCN (J = 3→2) profile indicates inflow motion.

Connection to larger scales

The hornlike H36α distribution and its broad line width [the Gaussian full-width at half-maximum (FWHM) is 393 ± 44 km s^{-1}; Fig. 3B] indicate a parsec-scale origin of the ionized outflow. The line profile of H36α is almost symmetric, indicating detection of both the receding and approaching sides of the ionization cone. By contrast, previous near-infrared coronal line profiles showed only blue tails, with velocities consistent with our H36α data (*27*), presumably because the receding side was obscured by dust in the earlier observations. The submillimeter H36α recombination line is much less affected by dust extinction, so it probes both sides of the cone. Using the previously constrained electron temperature of the circumnuclear region (*29*) and assuming local thermodynamic equilibrium, we estimated the volume-averaged electron density of this region to be ~7400 cm^{-3} (*29*). We measured the outflow velocity from the line profile (~200 km s^{-1}) and then used it to estimate the ionized outflow rate as ~0.04 $M_\odot$ year^{-1}. The kinetic power of this outflow (~5 × 10^{38} erg s^{-1}) corresponds to 0.001% of the AGN bolometric luminosity, which is considerably smaller than predicted

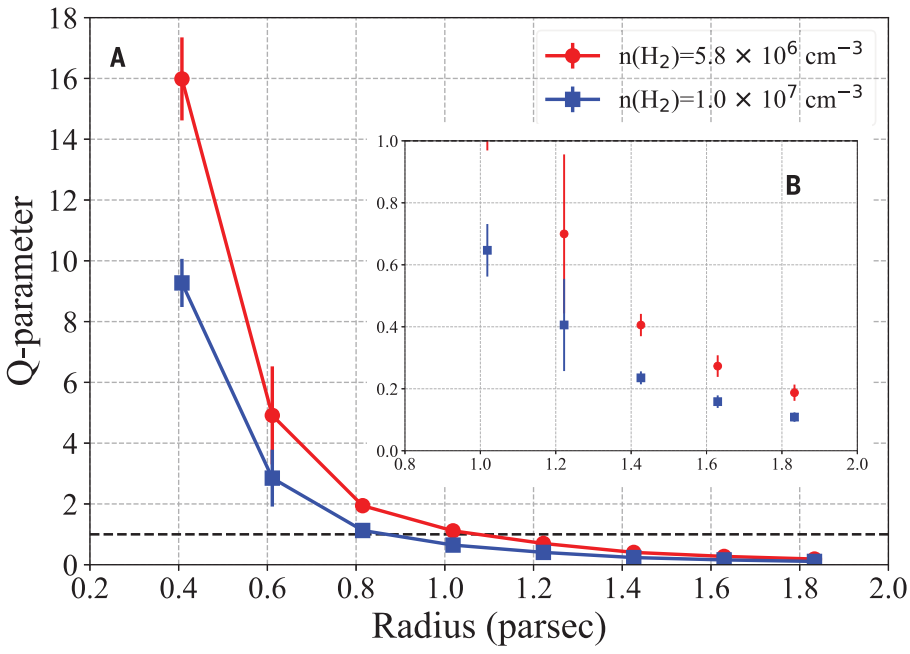


Fig. 4. Gravitational stability of the dense gas disk. (**A**) Radial profiles of the Toomre-*Q* parameter, plotted for two assumed values of the gas volume density (see legend). The horizontal dashed line indicates the critical value for stability, $Q = 1$. Error bars indicate 1σ uncertainty. (**B**) Same as panel A, but zoomed on the region $r > 0.8$ pc and $Q < 1$.

by some AGN-feedback models, e.g., ~5% in the energy-conserving feedback model (*20*).

Our measured inflow and outflow rates imply that only small fractions of the inflowing mass are consumed by black hole growth (< 3%) or expelled in the ionized gas outflow (< 20%). Therefore, the majority (>80%) of the inflowing mass must be converted to outflows in other gas phases. The two [C I] $^3P_1$–$^3P_0$ peaks in the minor-axis PVD (offset ~0.1 arcsec, equivalent to 2 pc), where CO ($J = 3\rightarrow2$) becomes fainter, are offset from the V_{sys} (446 km s^{-1}) by ~15 km s^{-1} (Fig. 3C). Considering the orientation of the Circinus Galaxy [southeast is the near side (*11*)], these offset components indicate outflows of atomic gas. The faint CO emission in that location might be caused by weak molecular outflows. The maximum [C I] $^3P_1$–$^3P_0$ outflow velocity (~40 km s^{-1}) is slower than the escape velocity from the SMBH gravitational potential at this scale (*29*). Therefore, we expected these outflows visible in [C I] $^3P_1$–$^3P_0$ to stall and become backflows to the CND, causing additional turbulence that could support the geometrical thickness of the atomic gas disk (*33*). Figure 3D summarizes the multiphase gas flows that we identified.

Physical process driving the accretion

To determine the physical process responsible for the accretion seen in the HCN ($J = 3\rightarrow2$) profile, we first considered turbulent viscosity in the dense gas disk. The turbulent accretion time scale is $t_{acc} = r^2/\nu_{vis}$, where ν_{vis} is the viscous parameter, defined as $\nu_{vis} = \alpha\,\sigma_{disp}h$, where h is the disk scale height and α is an efficiency parameter in the range $0 < \alpha < 1$ (*36*). This gives an accretion time scale of $1.8 \times 10^5 \times \alpha^{-1}$ year^{-1} at $r = 0.27$ pc (taking σ_{disp} ~ 15 km s^{-1} and $h \sim 0.1 \times r$, estimated from Fig. 2E). The enclosed dense gas mass within this radius is $\pi r^2 \times 2h \times 2n_{H2} \times m_p$ ~3500 to 6100 $M_\odot$, again assuming that n_{H2} spans the range of HCN ($J = 3\rightarrow2$) critical densities given above. The mass inflow rate expected under this viscous accretion is then ~α × (0.02 to 0.03) $M_\odot$ year^{-1}. Even for the highest possible value of $\alpha = 1$, this inflow rate is an order of magnitude smaller than that which we deduced from the infall model. We therefore found viscous accretion to be an unlikely mechanism at this scale of several parsecs.

Alternatively, the torques induced by nonaxisymmetric gravitational instabilities [caused by spiral arms or clumps (*37*)] or gas clump-clump collisions (*7*) are possible mechanisms for angular momentum transfer. We estimated the radial profile of the Toomre Q parameter, which expresses the balance between the self-gravity of the molecular gas and turbulent pressure (*38*), as $Q \equiv \kappa c_s/\pi G\Sigma$, where κ, c_s, G, and Σ are the epicyclic frequency of the orbit, the speed of sound, the gravitational constant, and the surface density of the gas mass, respectively. A geometrically thin disk becomes gravitationally unstable if $Q < 1$. We assumed $\Sigma = 2n_{H2} \times 2h \times m_p$ and the same range of n_{H2} as above. Substituting σ_{disp} into the expression for c_s and taking $h = (\sigma_{disp}/V_{rot}) \times r$, we obtained the profiles shown in Fig. 4. The derived values of Q are less than the critical value at $r > 1$ pc, indicating that the disk is gravitationally unstable, so we expected clump-clump collisions or instability-driven turbulence (*39*) to occur. For the regions slightly above $Q = 1$, we expected nonaxisymmetric structures (such as a nuclear spiral arm) to be induced, which would also cause a gravitational torque. Some nonaxisymmetric structures are visible in the HCN disk (Fig. 1C).

By contrast, the $Q > 1$ at $r < 1$ pc (Fig. 4). A dense gas mass of > 3 × 10^6 $M_\odot$ would be required to make the disk gravitationally unstable at $r < 0.4$ pc if the disk scale height is $h \sim 0.1 \times r$, as we estimated above. We ruled out such a high mass of dense gas on the basis of our HCN ($J = 3\rightarrow2$) rotation curve, which indicates that the enclosed mass is ~2 × 10^6 $M_\odot$ within $r = 1$ pc. Kinematic viscosity due to magnetohydrodynamic turbulence would not be sufficient; numerical models of that process predict $\alpha \sim 10^{-4}$ (*40*). To drive accretion on subparsec scales would require a further denser (>10^8 cm^{-3}) but thin disk, which might correspond to the disk seen in the H_2O maser observations (*24*). Our HCN ($J = 3\rightarrow2$) observations would not be sensitive to such a disk, but it is theoretically capable of driving the gravitational instability, so we prefer this explanation.

REFERENCES AND NOTES

1. D. M. Alexander, R. C. Hickox, *New Astron. Rev.* **56**, 93–121 (2012).
2. J. Kormendy, L. C. Ho, *Annu. Rev. Astron. Astrophys.* **51**, 511–653 (2013).
3. A. Soltan, *Mon. Not. R. Astron. Soc.* **200**, 115–122 (1982).
4. B. M. Peterson, *An Introduction to Active Galactic Nuclei* (Cambridge Univ. Press, 1997)
5. P. F. Hopkins, L. Hernquist, T. J. Cox, D. Keres, *Astrophys. J. Suppl. Ser.* **175**, 356–389 (2008).
6. M. J. Koss *et al.*, *Nature* **563**, 214–216 (2018).
7. I. Shlosman, M. C. Begelman, J. Frank, *Nature* **345**, 679–686 (1990).
8. P. F. Hopkins, L. Hernquist, *Astrophys. J. Suppl. Ser.* **166**, 1–36 (2006).
9. T. Storchi-Bergmann, A. Schnorr-Müller, *Nat. Astron.* **3**, 48–61 (2019).
10. F. Combes *et al.*, *Astron. Astrophys.* **565**, A97 (2014).
11. T. Izumi, K. Wada, R. Fukushige, S. Hamamura, K. Kohno, *Astrophys. J.* **867**, 48 (2018).
12. S. García-Burillo *et al.*, *Astron. Astrophys.* **652**, A98 (2021).
13. T. A. Davis *et al.*, *Mon. Not. R. Astron. Soc.* **496**, 4061–4078 (2020).
14. R. C. Levy *et al.*, *Astrophys. J.* **935**, 19 (2022).
15. C. M. V. Impellizzeri *et al.*, *Astrophys. J. Lett.* **884**, L28 (2019).
16. M. Imanishi *et al.*, *Astrophys. J.* **902**, 99 (2020).
17. G. R. Tremblay *et al.*, *Nature* **534**, 218–221 (2016).
18. T. Rose *et al.*, *Mon. Not. R. Astron. Soc.* **496**, 364–380 (2020).
19. T. Rose *et al.*, *Mon. Not. R. Astron. Soc.* **518**, 878–892 (2023).
20. A. King, K. Pounds, *Annu. Rev. Astron. Astrophys.* **53**, 115–154 (2015).
21. S. Veilleux, R. Maiolino, A. D. Bolatto, S. Aalto, *Astron. Astrophys. Rev.* **28**, 2 (2020).
22. R. Tully *et al.*, *Astron. J.* **138**, 323 (2009).
23. P. Arévalo *et al.*, *Astrophys. J.* **791**, 81 (2014).
24. L. J. Greenhill *et al.*, *Astrophys. J.* **590**, 162–173 (2003).
25. A. S. Wilson *et al.*, *Astrophys. J.* **120**, 1325–1341 (2000).
26. M. A. Fonseca-Faria, A. Rodríguez-Ardila, M. Contini, V. Reynaldi, *Mon. Not. R. Astron. Soc.* **506**, 3831–3852 (2021).
27. F. Müller Sánchez *et al.*, *Astron. Astrophys.* **454**, 481–492 (2006).
28. K. R. W. Tristram *et al.*, *Astron. Astrophys.* **664**, A142 (2022).

29. Materials and methods are available as supplementary materials.
30. P. R. Maloney, D. J. Hollenbach, A. G. G. M. Tielense, *Astrophys. J.* **466**, 561 (1996).
31. T. Izumi *et al.*, *Astrophys. J.* **898**, 75 (2020).
32. E. M. Di Teodoro, F. Fraternali, *Mon. Not. R. Astron. Soc.* **451**, 3021–3033 (2015).
33. K. Wada, M. Schartmann, R. Meijerink, *Astrophys. J. Lett.* **828**, L19 (2016).
34. J. Di Francesco, P. C. Myers, D. J. Wilner, N. Ohashi, D. Mardones, *Astrophys. J.* **562**, 770–789 (2001).
35. K. Wada, *Astrophys. J.* **559**, L41–L44 (2001).
36. N. I. Shakura, R. A. Sunyaev, *Astron. Astrophys.* **24**, 337–355 (1973).
37. P. Goldreich, D. Lynden-BellIII, *Mon. Not. R. Astron. Soc.* **130**, 125–158 (1965).
38. A. Toomre, *Astrophys. J.* **139**, 1217–1238 (1964).
39. K. Wada, G. Meurer, C. A. Norman, *Astrophys. J.* **577**, 197–205 (2002).
40. Y. Kudoh, K. Wada, C. Norman, *Astrophys. J.* **904**, 9 (2020).
41. T. Izumi *et al.*, Data for: Supermassive black hole feeding and feedback observed on subparsec scales, UTokyo Repository (2023); https://doi.org/10.15083/0002007911.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their thorough reading and constructive comments. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. **Funding:** This work was supported by the Japan Society for the Promotion of Science (JSPS) through KAKENHI (grants JP20K14531 and JP23K03462 to T.I.; grant JP21H04496 to K.W.; grant JP17H06130 to K.K.; grant JP21K03632 to M.I.; grant JP19K03937 to K.N.; grant JP20K14529 to T.K.; and grants JP20H00181, JP22H00158, and JP22H01268 to Y.F.); NAOJ ALMA (scientific research grants 2020-14A to Y.K. and K.W. and 2022-21A to T.I. and Y.F.); and by ALMA Japan Research Grant for the NAOJ ALMA (project code NAOJ-ALMA-271 to T.I.). **Author contributions:** T.I. led the project, analyzed the data, produced the figures, and led the manuscript writing. K.W., Y.K., and S.B. contributed to the theoretical interpretation and produced the figures. All authors contributed to scientific discussion and writing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The ALMA data are available from the ALMA archive at https://almascience.nao.ac.jp/aq/. This paper makes use of the following ALMA data: ADS/JAO. ALMA #2019.1.00014.S, #2017.1.00575.S, and #2018.1.00581.S. Our reduced ALMA data cubes, dynamical model input, and dynamical model results are archived at the UTokyo Repository (*41*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf0569
Materials and Methods
Figs. S1 to S9
Tables S1 to S2
References (*42*–*66*)

Submitted 27 September 2022; accepted 26 September 2023
10.1126/science.adf0569

IONTRONICS

Cascade-heterogated biphasic gel iontronics for electronic-to-multi-ionic signal transmission

Weipeng Chen[1,2]†, Linxin Zhai[3]†, Suli Zhang[4,5]†, Ziguang Zhao[1,2]†*, Yuhao Hu[1,2], Yun Xiang[1], Huirong Liu[4,5], Zhiping Xu[3], Lei Jiang[1,2], Liping Wen[1,2]*

Currently, electronics and iontronics in abiotic-biotic systems can only use electrons and single-species ions as unitary signal carriers. Thus, a mechanism of gating transmission for multiple biosignals in such devices is needed to match and modulate complex aqueous-phase biological systems. Here we report the use of cascade-heterogated biphasic gel iontronics to achieve diverse electronic-to-multi-ionic signal transmission. The cascade-heterogated property determined the transfer free energy barriers experienced by ions and ionic hydration-dehydration states under an electric potential field, fundamentally enhancing the distinction of cross-interface transmission between different ions by several orders of magnitude. Such heterogated or chemical-heterogated iontronics with programmable features can be coupled with multi-ion cross-interface mobilities for hierarchical and selective cross-stage signal transmission. We expect that such iontronics would be ideal candidates for a variety of biotechnology applications.

In biological systems, neural networks with complex morphologies and highly polarized interfacial architectures can support elaborate bioionic and biochemical signal communications between different neurons (*1*, *2*). Such structures and functions of complex neural networks provide inspiration for designing cascade-gated architectures capable of implementing neuron-like multivariate ionic or chemical signal transport that could interface with and modulate physiological processes.

Recently, artificial electronic and iontronic systems, which broke the information barriers between abiotic and biotic interfaces, have attracted considerable attention for application in biosensors (*3*–*6*), neuroprosthetics (*7*, *8*), and implantable neuromorphic devices (*9*–*12*). Electronics, especially neuromorphic electronics (*11*–*14*), directly use programmable electrical impulse signals to treat biological interfaces as signal transducers for modulating physiological activities (*14*–*16*), whereas iontronics exhibit the capability of individually performing the conversion between intrinsic ionic and electrical signals for abiotic-biotic systems (*17*, *18*). However, compared with biological neuronal networks with diverse biosignal transmission invoked by neural action potentials (fig. S1), most electronics and iontronics are still restricted by the single-species attributes of signal carriers (electrons or single ions) that cannot carry more biocompatible information. These existing iontronics constructed from conventional gated or nongated materials do not faithfully mimic multi-ionic cotransport, thereby limiting their characteristic signal expression in matching biological tissues (*19*). Ionic current signal transmission does not even discriminate between various ions in hydrogel-based iontronics (*20*–*22*). In the realm of iontronics, diverse electronic-to-multi-ionic transmission and processing in aqueous-phase biological environments remain a challenge.

Design and fabrication of cascade-heterogated biphasic gels

Here we report the use of a biphasic gel iontronic with cascaded heterointerface–gated properties to construct diverse ionic cross-medium transmission as a signal communication principle (Fig. 1A). The cascade-heterogated biphasic gel (HBG) has phase-separated heteronetwork structures that incorporate opposite binary gel phases: a microscale ion-enriched internal gel phase (IE phase) and a low-conductivity continuous gel phase (LC phase) (Fig. 1B). Specifically, the IE phase features hydrogel microinclusions that can store and transmit hydrated ions, and the LC phase serves as an organogel matrix that functions as the partially dehydrated ion transmission medium. Ionic cross-medium transmission can be triggered by external stimulation of both continuous- and pulse-based electrical signals, during which partial shedding and reconfiguration of cationic hydration shells, acting as ionic partial hydration and rehydration states (Fig. 1C), must occur alternately and consecutively. Thus, biphasic multi-interfaces induce cascade-heterogated effects that determine how ions experience transfer free energy barriers (E_a) and local driving force ($\propto \Delta V$) under an electric potential (Fig. 1, D and E). Because of the association of these factors with ionic hydration-dehydration energies, the distinction of cross-interface transmission between multivariate cations (X^+ and X^{n+}) is enhanced by several orders of magnitude. By way of the sorting and control of ionic transfer energy barriers at heterogated interfaces, the HBG system can process electronic-to-multi-ionic hierarchical transmission and selective ionic cross-stage transmission.

An in situ phase-separated polymerization strategy was introduced to fabricate HBG iontronics in which interfacial covalent interactions could maintain stable heterogated architectures (fig. S2). Heterogeneous structures of HBGs with different phase compositions were confirmed by confocal laser scanning microscopy (CLSM) (Fig. 1B and fig. S3) and scanning electron microscopy (SEM) (fig. S4). In CLSM images, it was observed that IE microinclusions (blue-stained) were uniformly dispersed in the LC phase (orange-stained) and found that their average diameters increased from 1.48 to 2.67 μm upon increasing the ratio of the IE phase (fig. S5). These results indicate that cascade-heterogated structures are modulated by the microphase separation degree. Ion species and concentrations did not affect HBG heterostructure (figs. S6 and S7). These samples were denoted as HBG-*y*, where *y* represents the volume ratio of the LC phase to the IE phase (fixed to 1). Desirable mechanical properties and environmental stability of HBG iontronics ensure high-reliability ion transport in multimedia-based environments (figs. S8 and S9).

Ionic transmission characteristics of cascade-heterogated biphasic gels

In nongated hydrogel iontronics, X^+ and X^{n+} cations are highly hydrated and can continuously move with extremely low viscous restrictions in water-rich environments and easily traverse the neutral hydrogel network in an electric potential field (Fig. 2A). There is no obvious differentiation between the ΔE of hydrated X^+ and X^{n+} cations during homogeneous transmission (fig. S10). In HBG systems, the cascade-heterogated mechanism can effectively govern ion transport owing to the ionic intrinsic ΔE under the external stimulation of continuous or pulse-based electrical signals (Fig. 2B). In the cascaded cross-interface ion transmission, the partial shedding and reconfiguration of cation hydration shells, which correspond to the ionic partial hydration and rehydration states, occurs alternately and successively. To further illustrate the heterogated effect derived from asymmetric chemical and spatial structures, we established the molecular dynamics model of the heterointerface

[1]CAS Key Laboratory of Bio-inspired Materials and Interfacial Science, Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, P. R. China. [2]School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, P. R. China. [3]Applied Mechanics Laboratory, Department of Engineering Mechanics, Tsinghua University, Beijing 100084, P. R. China. [4]Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Capital Medical University, Beijing 100069, P. R. China. [5]Beijing Key Laboratory of Metabolic Disorder Related Cardiovascular Disease, Capital Medical University, Beijing 100069, P. R. China.
***Corresponding author. Email: zhaoziguang@ucas.ac.cn (Z.Z.); wen@mail.ipc.ac.cn (L.W.)**
†These authors contributed equally to this work.

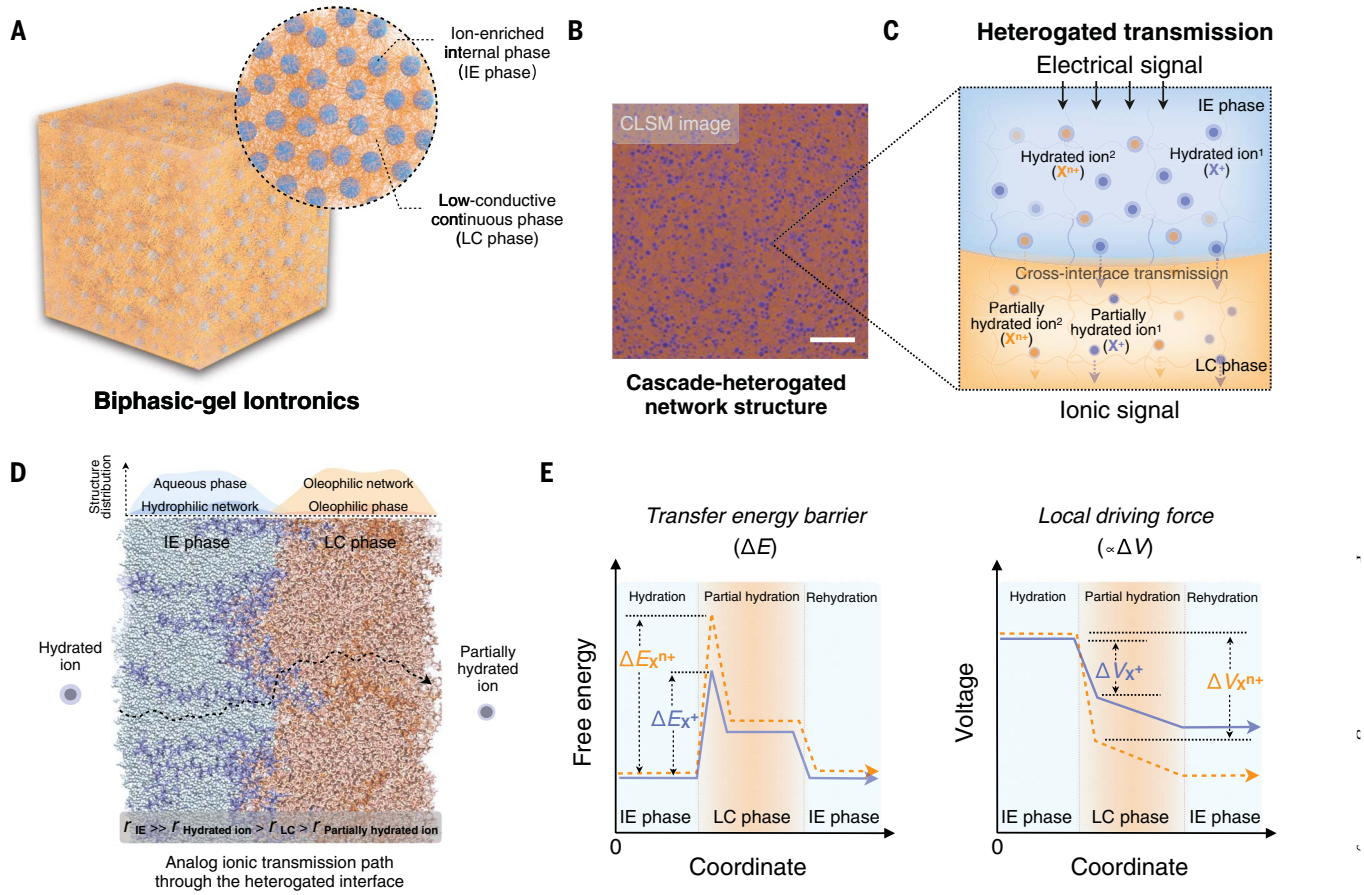


Fig. 1. Heterogeneous structures and cross-interface ion transmission of cascade-heterogated biphasic gel iontronics. (**A**) Structural illustration and (**B**) CLSM image of the HBGs with phase-separation heteronetworks. Scale bar, 20 μm. (**C**) Heterogated cross-interface ion transmission. Owing to the distinct hydration-dehydration energies between different X^+ and X^{n+}, the cascade-heterogated effects fundamentally enlarged the hierarchy discrepancy of the ionic transfer energy barriers to construct diverse electronic-to-multi-ionic signal transmission by the application of electrical signals. (**D**) Analysis of the network structure distribution at the heterogated interface and the analog ionic transmission path through the heterogated interface from the hydrated state in the IE phase to the partially hydrated state in the LC phase. (**E**) The obvious distinction of transfer energy barrier (ΔE) and the local driving forces ($\propto \Delta V$) of X^+ and X^{n+} featuring the hydration and partial hydration states across the heterogeneous interfaces.

to simulate the ionic transmission path and the related transfer free energy barrier (fig. S11). Crossing the heterointerface from the IE phase to the LC phase, the attractive and repulsive interactions between cation hydration shells and IE and LC phases, as well as the steric limitations of heterogated interfaces, led to a substantial increase in ΔE. Meanwhile, the resultant transport resistance could be counteracted by the local driving force from the electric field, which was reflected by the notable ΔV. Ionic transmission paths in the IE phase with an average estimated size of 6.60 to 7.26 nm were found to be larger than those of hydrated ions, and so a partial dehydration transformation reduced the effective size of hydrated ions, allowing them to access the LC phase, in which the estimated path size of 1.5 to 3.0 Å was larger than that of the partially hydrated ions (Fig. 1D and fig. S12A). The heterogated interface with a high degree of permeation also acted as a transition layer to reduce the driving force of ions across the heterointerface, and the interfacial gating thickness did not affect the ionic cross-interface energy consumption originating from its hydration-dehydration transformation (fig. S13). Subsequently, an energy compensation existed for the disengagement from the attractive interaction between the ion and the IE phase, and the partially hydrated ion with a related high transfer energy in the LC phase was relayed to the next IE microinclusions, accompanied by an energetically favorable reconfiguration to realize ion rehydration. For one unit of cascaded cross-interface ion transport (fig. S12B), the ratios of K^+, Ca^{2+}, and Fe^{3+} hydration-dehydration energies and ion mobilities between the biphasic system were simulated to evaluate different cross-interface ΔE values ($\Delta E_{Fe^{3+}} > \Delta E_{Ca^{2+}} > \Delta E_{K^+}$) of the three typical ions (fig. S12, C and D). Specifically, ionic radius, electrostatic force, hydrogen bonding, and other weak interactions, which together affect the stability of the ionic hydration shell, were believed to be intrinsic factors in determining the ionic hydration-dehydration energy (*23*, *24*). Theoretically, ions with high charge densities can lead to the strong electrostatic ordering of nearby water molecules to stabilize their hydration shells, thus necessitating the higher hydration-dehydration energies to overcome cross-interface transfer energy barriers (*25*).

Figure 2, C and D, and fig. S14 show similar ionic current variations of nongated hydrogels containing 1 mM K^+, Ca^{2+}, or Fe^{3+} under different voltage gradients, and their comparable transfer potentials were confirmed in terms of the corresponding voltage thresholds (all 1 mV). These indiscriminate results demonstrate that homogeneous networks of nongated hydrogels had difficulty recognizing ionic intrinsic features for the distinction of ion transport. In contrast, the ionic current properties of HBGs reflected obviously different

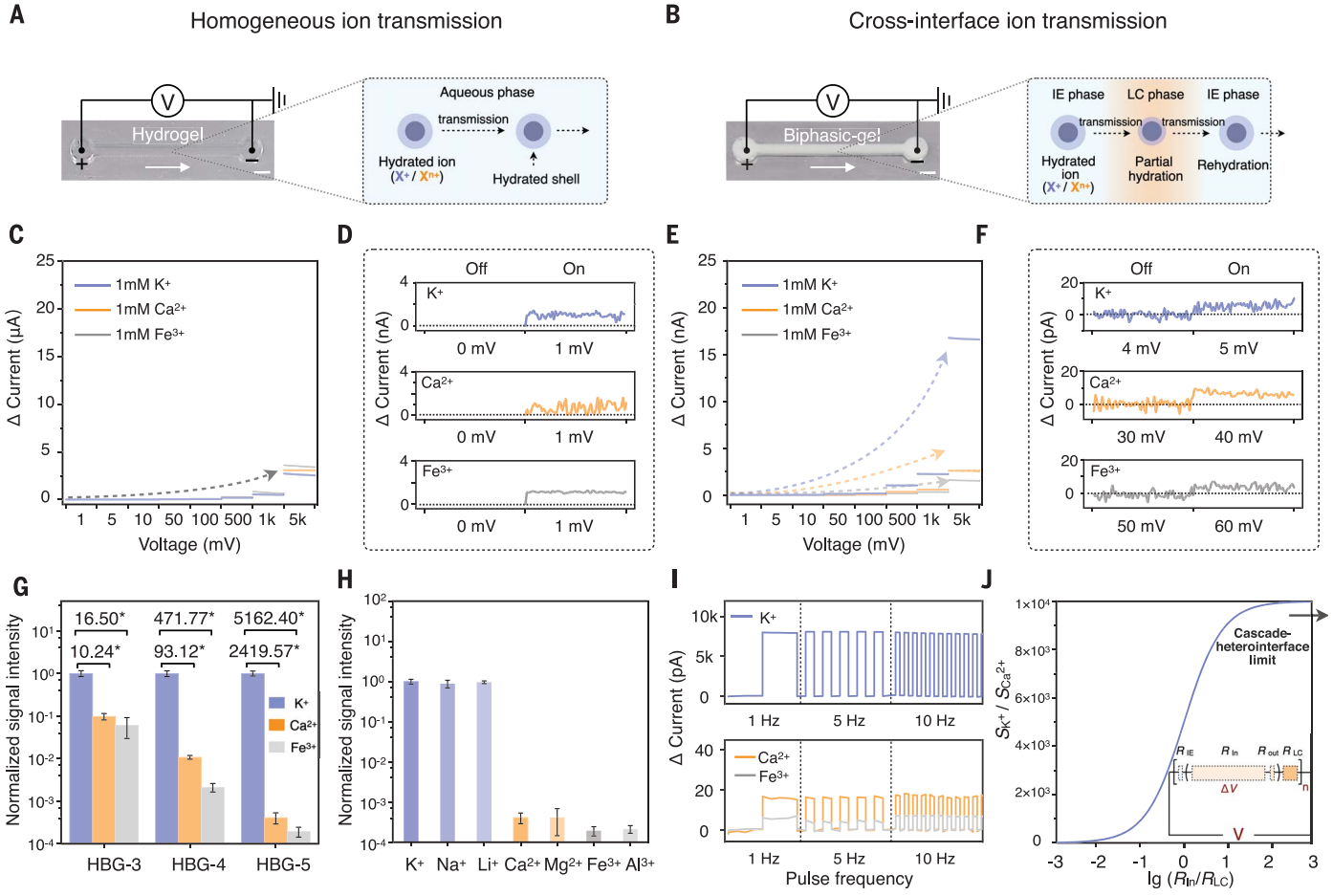


Fig. 2. Ionic transmission characteristics of nongated hydrogels and cascade-heterogated biphasic gels. (**A**) Homogeneous and (**B**) cross-interface ion transmission of the nongated hydrogel and HBGs, respectively. Scale bar, 3 mm. (**C** and **D**) Ionic current properties of the nongated hydrogels and (**E** and **F**) the HBGs containing 1 mM K^+, Ca^{2+}, or Fe^{3+} under varied applied voltages. (**G**) Statistically normalized ionic signal intensities of the HBGs with different phase compositions (data are mean ± SD, *N* = 3). (**H**) Normalized ionic signal intensities at 5 V of HBG-5 containing 1 mM of different ions. The ionic cross-interface transport capacity corresponds to the ionic hydration-dehydration energy (data are mean ± SD, *N* = 3). (**I**) Obvious distinction of the pulse-based K^+, Ca^{2+}, and Fe^{3+} ionic current properties of the HBG-5 under electrical signal stimulations with varying frequencies of 1, 5, and 10 Hz. (**J**) Calculated ionic signal intensity (*S*) ratio of K^+/Ca^{2+} in the cascade-heterointerface system with different cascade degrees. The inset shows the equivalent circuit of *n* cross-interface ionic transmission units.

trends under a range of varied applied voltages (Fig. 2, E and F). The voltage threshold of HBG-3 with 1 mM K^+ was 5 mV, and 40 and 60 mV, respectively, with Ca^{2+} and Fe^{3+}. To gain further insight into the relationship between the ionic cross-interface transmission and the phase-separation composition, ionic signal transmission properties of HBGs varying with cascade-heterogated effects were characterized. The ionic currents of the low-conductivity organogel separated from the biphasic gel precursors were initially measured. All samples were in a nonconductive state at 5 V, confirming that the ions were not stored and then continuously transported in the LC phase (fig. S15). The differences in ionic (K^+, Ca^{2+}, Fe^{3+}) signals derived from the heterojunction biphasic gel with a single and uniform heterointerface were still rough in comparison with HBGs featuring cascade-heterogated interfaces (figs. S16 and S17). By modulating cascade-heterogated effects of HBGs, the ratios of on-state transfer potentials and the normalized signal intensities between different cations exhibited increasing trends induced by increasing the degrees of microphase separation (Fig. 2G and figs. S18 and S19A). Typically, the ratios of voltage thresholds (K^+ to Fe^{3+} or K^+ to Ca^{2+}) of HBG-5 were 66.76 and 53.34. The normalized signal intensity ratios of K^+ to Fe^{3+} and K^+ to Ca^{2+} in HBG-5 reached 5162.40 and 2419.57. In contrast, the values of K^+ to Fe^{3+} and K^+ to Ca^{2+} in nongated hydrogels were only 2.19 and 1.74, respectively. Such homogeneous ion transmission depended on the ionic intrinsic transfer coefficients under an electric potential. The heterogated effect also held for iontronic systems in the presence of a high cation concentration up to 100 mM (fig. S20). Otherwise, the normalized signal intensities and related voltage thresholds of HBG-5 indicated that different ionic cross-interface transport capacities were related to the theoretical trends of their hydration-dehydration energies (Fig. 2H, fig. S19B, and table S1). For both isovalent and aliovalent ions, the ion with low hydration-dehydration energy corresponded to low cross-interface ΔE, thus exhibiting a relatively high ionic signal intensity and a low voltage threshold.

Compared with the homogeneous ion transmission of nongated hydrogels (fig. S21), cascade-heterogated effects also demonstrated an obvious distinction between pulse-based K^+, Ca^{2+}, and Fe^{3+} ionic current properties under electrical stimulations at frequencies from 0.1 to 100 Hz (Fig. 2I and fig. S22). The simulated relationship between ion dynamics and pulse-based frequencies (0.1 to 1000 Hz) at the heterogated interface was conducted, in which the simulated transmitted ion number and normalized simulated current were unrelated to the stimulus frequency (fig. S23). The duration of the high level in a single pulse cycle far exceeded the expected time

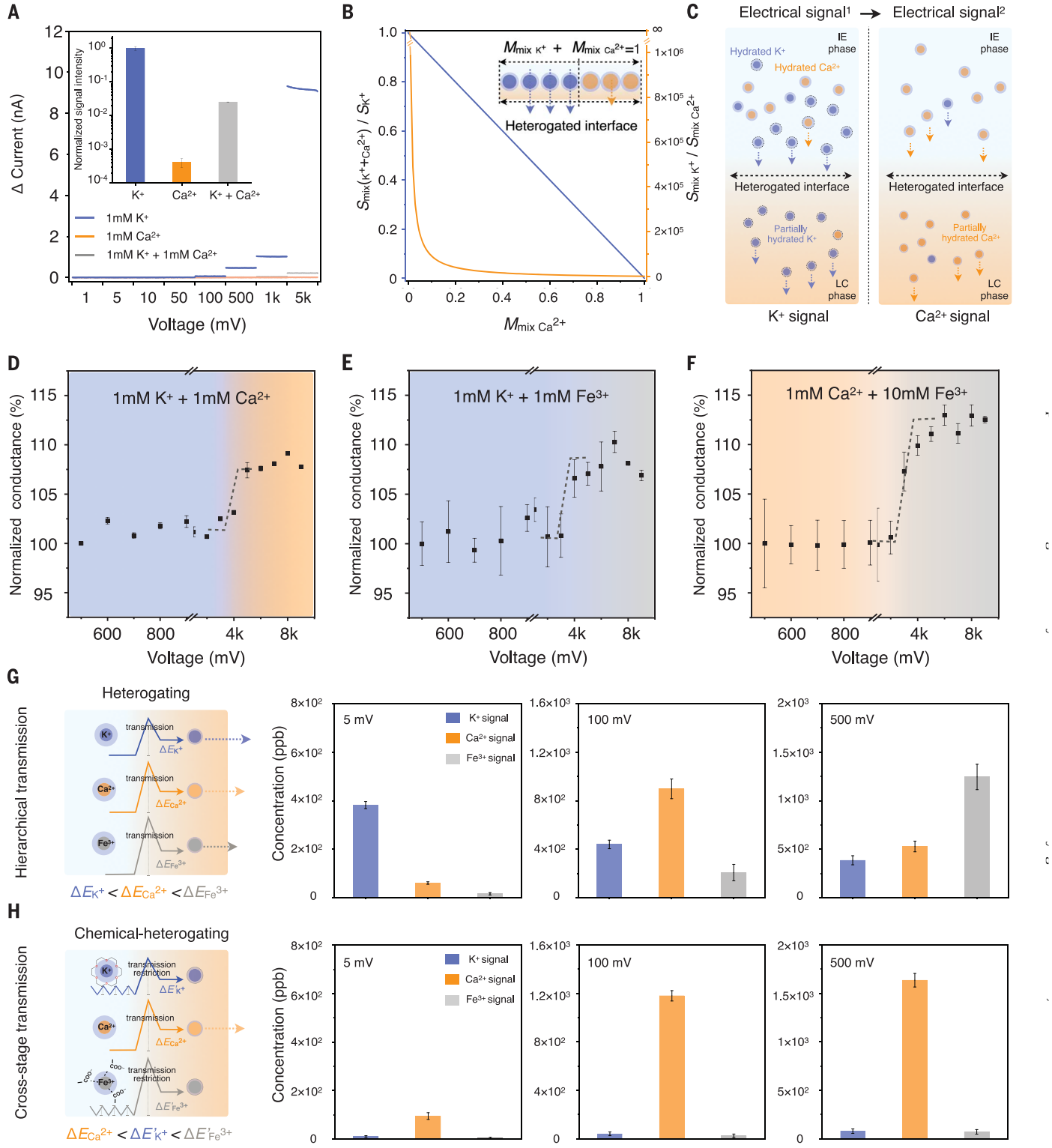


Fig. 3. Electronic-to-multi-ionic signal transmission of cascade-heterogated biphasic gels. (**A**) Ionic current characteristics of the HBG-5 containing different ions. The inset showed the corresponding normalized signal intensities at 5 V (data are mean ± SD, N = 3). (**B**) Simulated signal intensity ratios of the HBGs with different mixtures of K^+ and Ca^{2+} to that of the HBG with K^+ (blue line), which were attributed to the different occupation proportions of K^+ ($M_{mix\ K^+}$ as the K^+ molar concentration ratio) and Ca^{2+} ($M_{mix\ Ca^{2+}}$ as the Ca^{2+} molar concentration ratio) at the heterogated interface, and simulated independent signal intensity ratios of K^+ ($S_{mix\ K^+}$) to that of Ca^{2+} ($S_{mix\ Ca^{2+}}$) in the mixed system (orange line). (**C**) Schematic of preferential K^+ transport and hysteretic Ca^{2+} transport at the heterogated interface. (**D** to **F**) Normalized conductance of the HBG-5 containing different ions (data are mean ± SD, N = 3). (**G**) Hierarchical ionic signal transmission. The pulse-transmitted ionic signals of the cascade-heterogated iontronics were successively dominated by K^+, Ca^{2+}, and Fe^{3+} against different electrical pulse signals from 5 to 500 mV (data are mean ± SD, N = 3). (**H**) Selective Ca^{2+} cross-stage transmission derived from the chemical-heterogated iontronics (data are mean ± SD, N = 3).

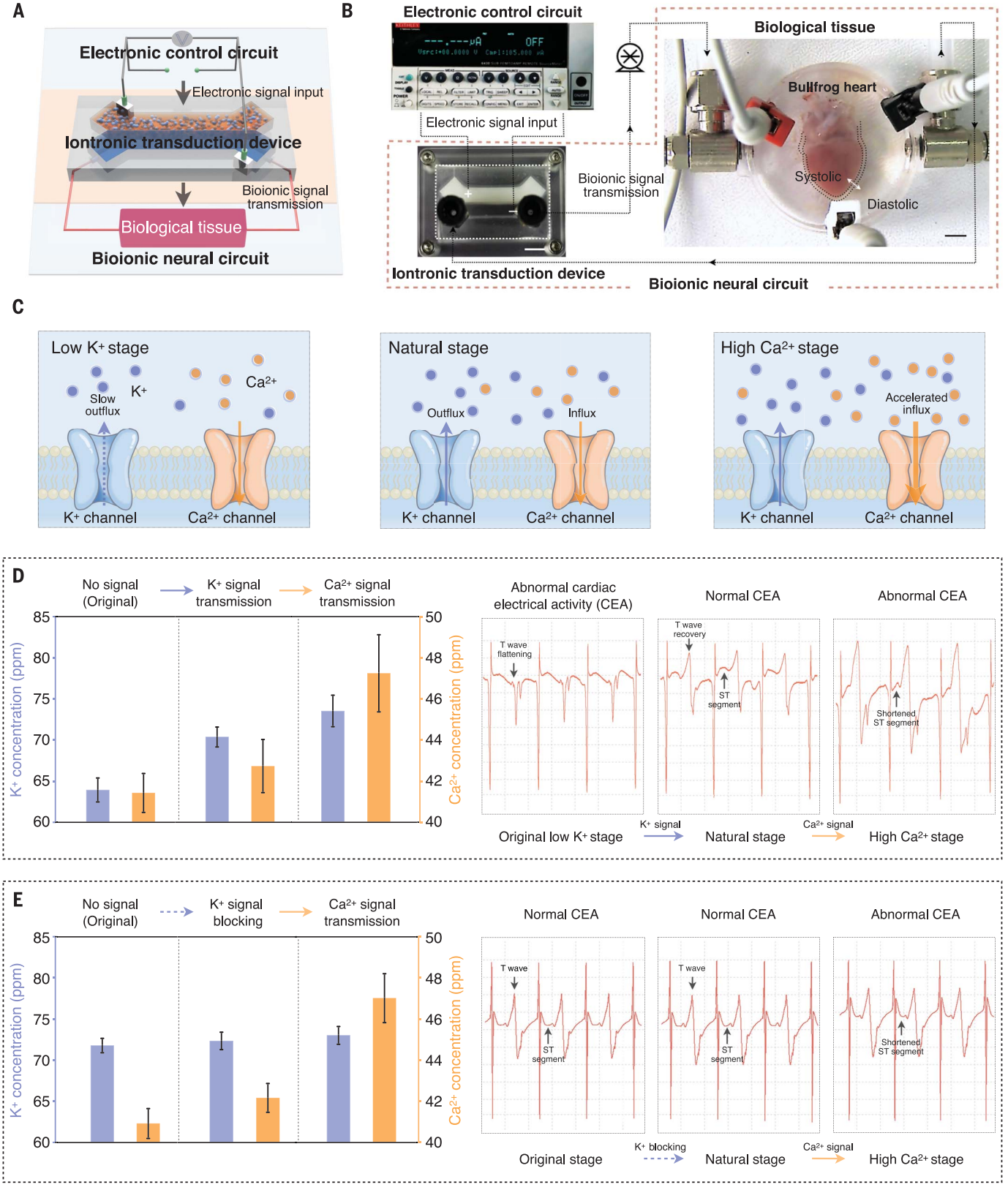


Fig. 4. Bioionic neurohumoral modulation. (**A**) Conceptual schematic of the iontronic signal transduction from electronic control circuits to bioionic neural circuits. (**B**) In an abiotic-biotic system, the cardiac electrical activities of a bullfrog heart were modulated by the bioionic neurohumoral signals from the iontronic transduction devices. Scale bars, 5 mm. (**C**) Exogenous bioionic neurohumoral signals modulate the permeability of cardiomyocytes and consequently alter the K^+ outflux and the Ca^{2+} influx in the cell membrane. (**D** and **E**) Bioionic neurohumoral modulation in the cardiac electrical activities of bullfrog hearts based on the cascade-heterogated (D) or chemical-heterogated (E) iontronic transduction devices (data are mean ± SD, N = 3).

for the cross-interface transmission of an ion, indicating that the cross-interface ion transport of HBG systems would be frequency-independent (1 to 1000 Hz).

To model cascade-heterogating effects, a series circuit model based on the nonlinear Poisson-Boltzmann theory was constructed to illustrate the relationship between cascaded heterointerfaces and K^+/Ca^{2+} transport distinction (Fig. 2J) (*26*, *27*). The total energy consumption ($\propto R_{Tot}$) can be divided into n units. For one cross-interface unit, the energy consumption can be divided into four sections: the consumption of transport in the IE phase (R_{IE}), across the heterogated interface from the IE phase to the LC phase (R_{In}), in the LC phase (R_{LC}), and across the heterogated interface from the LC phase to the IE phase (R_{Out}) (fig. S12B). R_{IE} was negligible owing to the weak resistance in the IE phase; and R_{Out} was also ignored because the energetically favorable reconfiguration for rehydration. On the basis of different cross-interface ΔE values, the ionic signal intensity (S) ratio of K^+/Ca^{2+} (which was equivalent to the total transport resistance ratio of K^+ to Ca^{2+}) was expressed as

$$S_{K^+}/S_{Ca^{2+}}(n) = 2\frac{nR_{LC}(Ca^{2+}) + nR_{In}(Ca^{2+})}{nR_{LC}(K^+) + nR_{In}(K^+)} = 2\frac{\frac{R_{LC}(Ca^{2+})}{R_{LC}(K^+)} + N\frac{R_{In}(Ca^{2+})}{R_{In}(K^+)}}{1+N} \quad (1)$$

where $N = \frac{R_{In}(K^+)}{R_{LC}(K^+)}$, and this N was used to compare the influence between the heterogated interface and the LC phase on the ion transport. When $N \to \infty$, the system could be regarded as the ideal cascade-heterointerface limit (denoted as the ideal HBG system)

$$S_{K^+}/S_{Ca^{2+}}(\text{Cascade} - \text{heterointerface limit}) = 2\frac{R_{In}(Ca^{2+})}{R_{In}(K^+)} = 2\exp\left[\frac{\Delta E_{In}(Ca^{2+}) - \Delta E_{In}(K^+) - 2e\Delta V_{In}(Ca^{2+}) + e\Delta V_{In}(K^+)}{k_B T}\right] \quad (2)$$

In contrast, when $N \to 0$, the system could be regarded as a single interface (denoted as a heterojunction)

$$S_{K^+}/S_{Ca^{2+}}(\text{Single interface}) = 2\frac{R_{LC}(Ca^{2+})}{R_{LC}(K^+)} \quad (3)$$

In practical conditions, the heterointerface density ($\propto N$) is often limited, and so the ionic signal intensity ratio of the HBG system cannot reach the theoretical limit. However, the cascade-heterogated interface of iontronics was sufficiently dense to ensure a high $S_{K^+}/S_{Ca^{2+}}$.

Electronic-to-multi-ionic signal transmission of cascade-heterogated biphasic gels

Figure 3A shows the anomalous decrease in the ionic current and the normalized signal intensity of HBG-5 containing a mixture of 1 mM K^+ and 1 mM Ca^{2+}. The presence of Ca^{2+} increased the total ion concentration, but the normalized signal intensity was only 0.025, which was one-fortieth that with K^+ alone. This result can be attributed to the fact that K^+ has a lower hydration-dehydration energy and a smaller ion-hydration size than Ca^{2+}, and so it was transported more favorably across the heterogated interface to consecutively undergo partial dehydration and rehydration under the stimulation of an electrical signal, resulting in relatively efficient ion transmission. However, the Ca^{2+} cross-interface mobility required a greater transfer energy, which reduced its own transport efficiency and formed interfacial occupations, causing steric hindrance and electrostatic repulsion at the heterogated interface, thereby further hindering K^+ and Ca^{2+} transport. Thus, the simulated signal intensity ratio of HBGs with different mixtures of K^+ and Ca^{2+} to that of HBGs with only the same concentration of K^+ was negatively linearly correlated with the Ca^{2+} occupation ratio ($M_{mix\ Ca^{2+}}$) at the heterogated interface (Fig. 3B, blue line). However, despite the decrease of ionic transmission paths resulting from such occupation, the simulated independent signal intensity ratio of K^+/Ca^{2+} in the mixed system remained at greater than three orders of magnitude (as $M_{mix\ Ca^{2+}} \leq 0.91$) (orange line). Figure 3C shows preferential K^+ transport and hysteretic Ca^{2+} transport ($\Delta E_{K^+} < \Delta E_{Ca^{2+}}$) at the heterointerface under segmental electrical signal stimulations. The Ca^{2+} occupation blocked the ion paths of the heterogated interface, also resulting in a relatively low ionic conductance below ~3 V. As the applied electrical signal increased, the normalized conductance curve suddenly increased, which was mainly attributed to the increased transport of hysteretic Ca^{2+} (Fig. 3D). This effect was also confirmed in HBGs containing different binary ionic mixtures (Fig. 3, E and F, and figs. S25 and S26).

Such cascade-heterogatings can fundamentally enlarge the hierarchy discrepancy of multi-ion transfer energy barriers for hierarchical signal transmission. The real-time variations in the transmitted ionic signal strengths derived from HBGs incorporated with ternary ionic mixtures were monitored by inductively coupled plasma mass spectrometry (fig. S27). Figure 3G and fig. S28 show that pulse-transmitted ionic signal strengths were successively dominated by K^+, Ca^{2+}, and Fe^{3+} against different input voltage signals ranging from 5 to 500 mV. In the absence of an input voltage signal, the HBGs maintained stable ion storage and avoided ion free diffusion to the surrounding aqueous phase environment. By applying a 5-mV pulse-based electrical signal with a frequency of 5 Hz, the transmitted K^+ signal of the HBG was triggered, and the transport of Ca^{2+} and Fe^{3+} remained hysteretic. After 2 hours, the K^+ signal strength increased from 6 to 400 parts per billion. Subsequently, by increasing the applied voltages to 100 and 500 mV, Ca^{2+}- and Fe^{3+}-dominated signals were observed, respectively. The cascade-heterogated effect allowed for the preferential transport of multiple ions by the sorting of ionic transfer energy barriers. Meanwhile, such transmitted ionic signal strengths of HBGs have already achieved levels compatible with those of specific physiological bioactivities (*28*, *29*). Unlike ion-selective membranes and hydrogels, our HBGs can store a stable ion source on demand and efficiently process intrinsic ionic signals to the ambient aqueous environment. The hydrogel iontronics inevitably led to uncontrolled ionic diffusion into the external environment, making it impossible to modulate or regulate ion transmission (fig. S29).

Considering the advantageous transport order of one type of ion under specific conditions, the hierarchical ion transmission could be regarded as a form of ion selectivity. We also designed a chemical-heterogated architecture to further regulate the ionic transfer energy barrier hierarchy for ion-selective cross-stage transmission. Owing to the coordination effects of specific ligand groups (18-crown-6 to K^+ and carboxylic group to Fe^{3+}) as chemical assistance in the IE phase (fig. S30A), ionic migration behaviors tended to be hopping between ligands to overcome dissociation energy barriers (*30*, *31*). Thus, the corresponding total energy consumption ($\propto R_{Tot}$) continued to increase, leading to a higher energy consumption in the IE phase than across the heterogated interface and breaking the inherence in ionic transport hierarchies that originated from their intrinsic hydration-dehydration energies. Such cascaded chemical-heterogated effects rendered the total equivalent transfer energy barriers of K^+ and Fe^{3+} significantly greater than that of Ca^{2+}, resulting in Ca^{2+} cross-stage signal and transmission restrictions of the other signal (Fig. 3H and fig. S30B). In contrast, dual-ligand (18-crown-6 and carboxylic group) hydrogel iontronics had no Ca^{2+} selectivity (fig. S31). In the homogeneous hydrogel network, chemical ligand effects were unable to further establish ion-scale confined-gating properties that governed ion transport.

Bioionic neurohumoral modulation

Neuronal signaling relies on the controlled transmission of various neurotransmitter molecules and ions (*32*, *33*). For iontronics, the aim is to regulate multiple biofunctional ionic signal transductions to interface with and control physiological activities in aqueous-based bioenvironments (*34*). To that end, the cascade-heterogated iontronic transduction device developed in this study to convert electronic input signals into diverse bioionic signals demonstrates its potential for application as a biocommunication carrier (Fig. 4, A and B). Meanwhile,

the good biocompatibility of the developed system was also demonstrated (figs. S32 and S33). Figure 4, C and D, outlines the transduction of biofunctional neurohumoral signals by cascade-heterogated and chemical-heterogated devices to regulate the permeability of cardiomyocytes and modulate cardiac electrical activities in bullfrog hearts. These results were confirmed by the alterations in the electrocardiogram (figs. S34 and S35 and movies S1 and S2). For abiotic-biotic systems, we anticipate that such iontronics could act as bridges for biocompatible signal processing and transmission between electronic control circuits and bioionic neural circuits.

Conclusions

Here we reported cascade-heterogated iontronics for diverse ionic signal transmission. These gels integrated opposite binary phase structures to form cascaded ionic transfer free energy barriers that were highly correlated with the hydration-dehydration energies of different ions. More impressively, cascade-heterogated interfaces fundamentally enlarged and controlled the hierarchical discrepancy of ionic transfer energy barriers to achieve multi-ionic hierarchical transmission and ion-selective cross-stage transmission. Until now, ion-gating has been primarily manifested in one- or two-dimensional systems; in this work, we developed a cascading strategy to expand the ion-gating mechanism to three-dimensional intronic systems for electronic-to-multi-ion transmission. Most iontronic and electronic devices employing conventional ion-gating or nongating struggle to effectively process multi-ion signal carriers in their potential applications (tables S2 and S3); however, these properties are critical for interfacing with and matching complex biological systems. HBG-based iontronics offer not only multispecies bioionic signals that are highly compatible with aqueous-phase biological systems but also advance programmable signal transmission functionalities that facilitate the parallel and multiplexed transmission of various bio-signals in abiotic-biotic systems. These cascade-heterogated iontronics serve as diverse platforms for possible extensions to more-biofunctional carriers and other signal transduction mechanisms. Meanwhile, it would be worthwhile to study the weight and coupling features of multispecies signal interchange in these cascade-gating systems. Despite the long road ahead for the construction of high-performance artificial neural networks, we expect that such cascade-heterogated networks will be applicable in hardware implementation for neuromorphic gating features, while also providing more signal morphologies and processing.

REFERENCES AND NOTES

1. J. Chen, Y. Kanai, N. J. Cowan, N. Hirokawa, *Nature* **360**, 674–677 (1992).
2. T. Yoshimura *et al*., *Cell* **120**, 137–149 (2005).
3. X. Xia, C. M. Spadaccini, J. R. Greer, *Nat. Rev. Mater.* **7**, 683–701 (2022).
4. Y. Liu *et al*., *Nat. Biotechnol.* **38**, 1031–1036 (2020).
5. S. Zhao *et al*., *Adv. Mater.* **30**, e1800598 (2018).
6. G. T. Go, Y. Lee, D. G. Seo, T. W. Lee, *Adv. Mater.* **34**, e2201864 (2022).
7. Y. Kim *et al*., *Science* **360**, 998–1003 (2018).
8. Z. Nie, J. W. Kwak, M. Han, J. A. Rogers, *Adv. Mater.* 10.1002/adma.202205609 (2022).
9. Y. Lee *et al*., *Nat. Biomed. Eng.* **7**, 511–519 (2023).
10. P. C. Harikesh *et al*., *Nat. Mater.* **22**, 242–248 (2023).
11. T. Sarkar *et al*., *Nat. Electron.* **5**, 774–783 (2022).
12. T. Wang *et al*., *Nat. Electron.* **5**, 586–595 (2022).
13. W. Wang *et al*., *Science* **380**, 735–742 (2023).
14. S. Chun *et al*., *Nat. Electron.* **4**, 429–438 (2021).
15. S. H. Jo *et al*., *Nano Lett.* **10**, 1297–1301 (2010).
16. L. Q. Zhu, C. J. Wan, L. Q. Guo, Y. Shi, Q. Wan, *Nat. Commun.* **5**, 3158 (2014).
17. H. Wei *et al*., *Nat. Commun.* **12**, 1068 (2021).
18. P. C. Harikesh *et al*., *Nat. Commun.* **13**, 901 (2022).
19. C. E. G. Hoskin, V. R. Schild, J. Vinals, H. Bayley, *Nat. Chem.* **14**, 650–657 (2022).
20. Y. Dobashi *et al*., *Science* **376**, 502–507 (2022).
21. P. Wang, G. Li, W. Yu, C. Meng, S. Guo, *Adv. Electron. Mater.* **8**, 2101269 (2022).
22. X. Liu, J. Liu, S. Lin, X. Zhao, *Mater. Today* **36**, 102–124 (2020).
23. R. Epsztein, R. M. DuChanois, C. L. Ritt, A. Noy, M. Elimelech, *Nat. Nanotechnol.* **15**, 426–436 (2020).
24. C. Lu *et al*., *J. Am. Chem. Soc.* **143**, 14242–14252 (2021).
25. P. A. Fernandes, M. N. D. S. Cordeiro, J. A. N. F. Gomes, *J. Phys. Chem. B* **104**, 2278–2286 (2000).
26. T. S. Badessa, V. A. Shaposhnik, *Electrochim. Acta* **231**, 453–459 (2017).
27. N. Rajeswari, S. Selvasekarapandian, M. Prabu, S. Karthikeyan, C. Sanjeeviraja, *Bull. Mater. Sci.* **36**, 333–339 (2013).
28. E. K. Hoffmann, I. H. Lambert, S. F. Pedersen, *Physiol. Rev.* **89**, 193–277 (2009).
29. M. Simard, M. Nedergaard, *Neuroscience* **129**, 877–896 (2004).
30. W. Wei, C. Xu, J. Ren, B. Xu, X. Qu, *Chem. Commun.* **48**, 1284–1286 (2012).
31. P. Lin, S. Ma, X. Wang, F. Zhou, *Adv. Mater.* **27**, 2054–2059 (2015).
32. S. Patel, R. Docampo, *Trends Cell Biol.* **20**, 277–286 (2010).
33. S. T. Keene *et al*., *Nat. Mater.* **19**, 969–973 (2020).
34. Y. Hou, X. Hou, *Science* **373**, 628–629 (2021).
35. W. Chen *et al*., Data for: Cascade-heterogated biphasic-gel iontronics for electronic-to-multi-ionic signal transmission processing, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.8319623.

ACKNOWLEDGMENTS

All animals used in the experiment were approved by the Animal Protection Ethics Committee of Capital Medical University (Ethics Number: AEEI-2014-062). We are grateful to X. Dong (Experimental Center for basic medical teaching, Basic Medical Sciences, Capital Medical University) for assistance with the biological experiments, and K. Wang (Capital Medical University) for helpful discussions. **Funding:** This research was supported by the National Natural Science Foundation (21988102, 22275183, 22122207, 11825203), the National Key R&D Program of China (2022YFB3805904, 2022YFB3805900), the CAS Project for Young Scientists in Basic Research (YSBR-039), and the China Postdoctoral Science Foundation (2022M713226). **Author contributions:** W.C., L.Z., S.Z., and Z.Z. contributed equally to this work. Z.Z., L.W., and W.C. proposed the research direction. Z.Z., L.W., and L.J. guided the project. W.C. and Z.Z. co-designed and performed the experiments. L.Z. and Z.X. performed the theoretical calculations. S.Z. and H.L. performed the biology experiments. Z.Z. and W.C. drafted the manuscript. All authors joined the data discussion and revised the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the main text or the supplementary materials. The data that support the findings of this study are available at Zenodo (*35*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg0059
Materials and Methods
Supplementary Text
Figs. S1 to S35
Tables S1 to S4
References (*36*–*72*)
Movies S1 and S2

Submitted 26 November 2022; resubmitted 22 July 2023
Accepted 22 September 2023
10.1126/science.adg0059

NEUROSCIENCE

Volitional activation of remote place representations with a hippocampal brain–machine interface

Chongxi Lai[1]*†‡, Shinsuke Tanaka[1]†‡, Timothy D. Harris[1], Albert K. Lee[1,2]*

The hippocampus is critical for recollecting and imagining experiences. This is believed to involve voluntarily drawing from hippocampal memory representations of people, events, and places, including maplike representations of familiar environments. However, whether representations in such "cognitive maps" can be volitionally accessed is unknown. We developed a brain–machine interface to test whether rats can do so by controlling their hippocampal activity in a flexible, goal-directed, and model-based manner. We found that rats can efficiently navigate or direct objects to arbitrary goal locations within a virtual reality arena solely by activating and sustaining appropriate hippocampal representations of remote places. This provides insight into the mechanisms underlying episodic memory recall, mental simulation and planning, and imagination and opens up possibilities for high-level neural prosthetics that use hippocampal representations.

The ability to simulate scenarios in one's mind is a hallmark of intelligence, as it facilitates the evaluation of past experiences and future plans. For instance, we can imagine walking around our previous workplace, or imagine how our current workplace might function if we rearranged the furniture. Such imagination requires an internal world model that can be flexibly accessed to construct possible scenarios (*1*–*3*).

The hippocampus is a brain region that is critical for memory and imagination (*1*, *4*–*6*). It holds a model of the environment (also called a cognitive map) (*7*, *8*) that could potentially be mentally traversed for the purpose of recall or simulation. In particular, the hippocampus contains spatial maplike representations of previously explored environments. Each environment's representation consists of place cells—neurons that fire selectively whenever an animal moves through specific locations (called the "place fields" of those cells) in that environment (*9*, *10*). This selective firing results in a distinct multicell activity pattern at each location in the environment, which, during physical navigation, can be used to decode the animal's current location from the ongoing pattern of neural activity (*11*). In contrast, a key aspect of imagination is the activation of neural representations that deviate from current sensory input, such as those that are nonlocal (i.e., represent locations away from one's current location). Previous work has shown brief and intermittent activation of nonlocal hippocampal spatial representations suggestive of the planning of specific paths within a cognitive map (*12*–*21*). However, it is unknown whether this activity is volitionally controlled or rather reflects passive memory-related processes that are presumably nonvolitional (*22*, *23*).

To test whether an animal can directly control its hippocampal activity according to its model of the world, we used a brain–machine interface (BMI) approach because, unlike with humans, we cannot simply ask animals to imagine scenarios. With BMI methods, we could reward animals for generating neural activity resembling the simulation of specific scenarios. More precisely, we could reward them for the volitional activation of specific nonlocal representations from the cognitive map—a fundamental building block of scenario simulation. BMI research has a rich history of directly testing for volitional control of activity patterns of neuronal ensembles in the motor cortex and related areas (*24*–*35*). In the hippocampus, it has been shown that the activity level of individual neurons (*36*, *37*) or the population activity related to individual stimuli (*38*) can be controlled. However, a real-time BMI that allows humans or animals to control their hippocampal population activity in terms of the content of their cognitive map (e.g., location representations) has never been demonstrated.

A hippocampal map–based BMI

We designed a real-time hippocampal BMI and two BMI tasks to investigate whether rats could navigate to goals ("jumper" navigation task), or move external objects to goals while remaining stationary ("Jedi" object location control task), within an immersive virtual reality (VR) environment solely by controlling the activity of a population of place cells. Each jumper or Jedi BMI experiment consisted of three phases (Fig. 1A). In phase 1, rats ran to a succession of arbitrary locations marked by a tall, visible goal cue placed in a familiar two-dimensional virtual arena ("running" task). Upon reaching each cue, liquid reward was delivered, the trial ended, and the cue moved to another location for the next trial. Animals were secured in a harness and could freely rotate their body and head direction on top of a spherical treadmill (*39*) while hippocampal CA1 neural activity was recorded (Fig. 1B, fig. S1, and movie S1). We applied a recently developed field-programmable gate array (FPGA)–based neural signal processor to perform low-latency (1 ms) assignment of extracellular spikes (recorded from 128 channels) to a population of hippocampal units (*40*, *41*). In the running task, treadmill movement updated the animal's location in the virtual environment, and many hippocampal units (i.e., place units) displayed spatially modulated activity (*39*, *42*–*44*) (Fig. 1B, blue arrows) similar to that in real-world environments (*8*–*11*). In phase 2, the binned spike counts from the most recent 1.5 or 5 s of activity of these place units and the animal trajectory from the running task were used to train a decoder (Fig. 1B, green arrows) that estimates the animal's current location from the neural data every 100 ms. We used a deep neural network for decoding (fig. S2), allowing the use of data augmentation for training—a method that improves both the decoder's performance given limited data and its noise robustness. In phase 3, the treadmill was disconnected from the VR system, and the animal's ability to control its own or an object's translational movement was limited to controlling its hippocampal activity, which was converted by the decoder into a specific location output every 100 ms (Fig. 1C). Note that the decoder was trained to estimate the animal's current location in the running task only, not its location in the subsequent BMI tasks, but, during BMI periods, the animal needed to generate activity corresponding to locations away from its current location.

BMI navigation task

In the jumper task, we tested whether animals could navigate to arbitrary goal locations as in the running task, except here by means of BMI-based first-person teleportation. After rats performed the running task for ~40 min (~120 trials) (Fig. 2, A and B, and movie S1), the data were used to train the decoder, which accurately estimated the rat's current location in the running task [validation set coefficient of determination (R^2) = 0.78 to 0.88] (Fig. 2C). Jumper trials were identical to running trials, except the animal's location was updated to the BMI-decoded location (smoothed with a 3-s sliding window to help reduce potential high-frequency visual jitter of the VR updates) (Fig. 2, D and E, and movie S1). If an animal did not reach the goal within 62 s, the trial

[1]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. [2]Howard Hughes Medical Institute and Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA.
*Corresponding author. Email: chongxi.lai@gmail.com (C.L.); alee31@bidmc.harvard.edu (A.K.L.)
†Present address: Howard Hughes Medical Institute and Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA.
‡These authors contributed equally to this work.

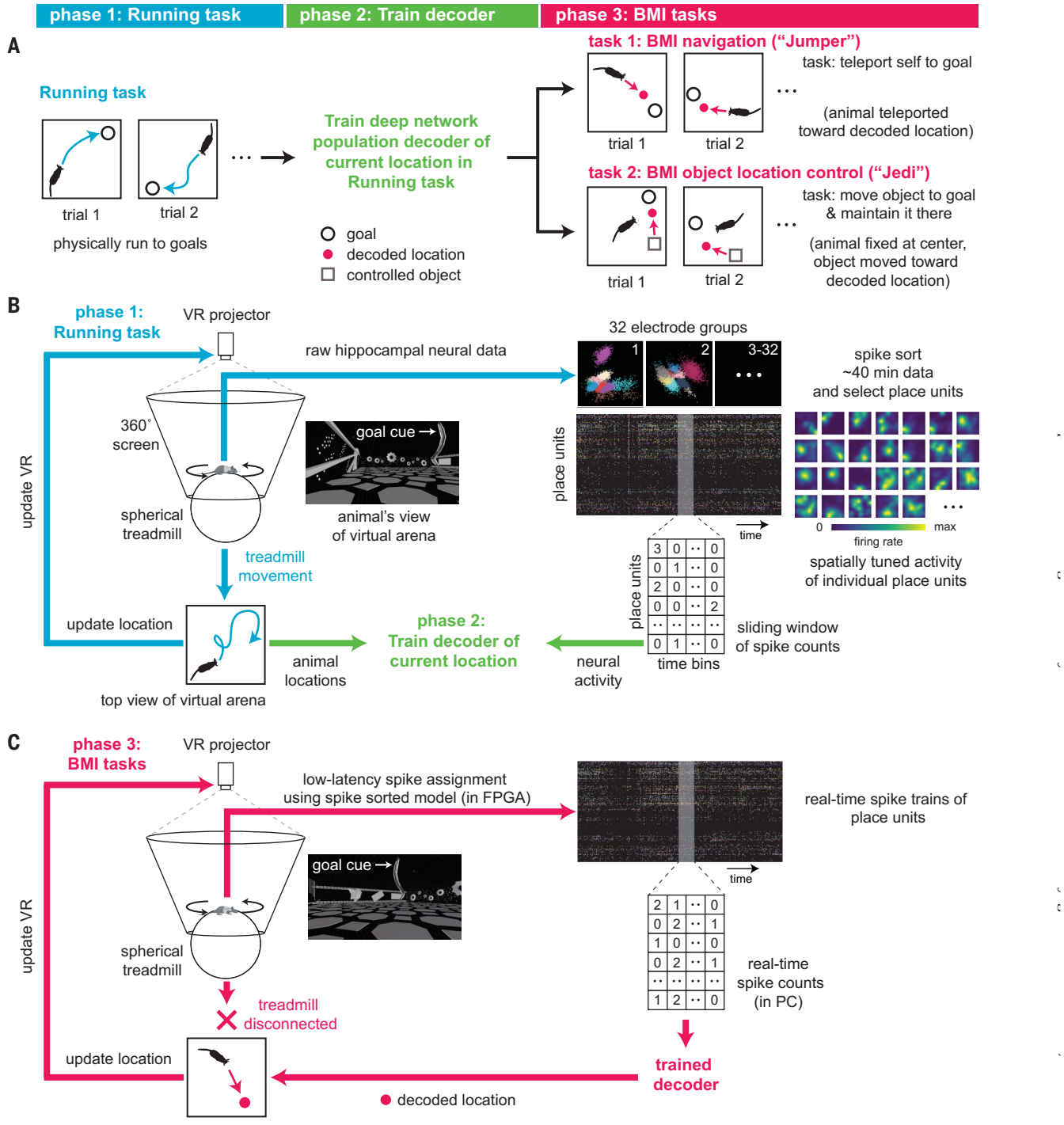


Fig. 1. Hippocampal map–based brain–machine interface in a virtual reality system. (**A**) Steps for performing the two different BMI experiments in this study. Rats first physically ran to a series of goals (running task), while their hippocampal neural activity and (virtual) location in a square arena were recorded. This data was used to train a decoder to take neural activity as input and output the animal's current location in the running task. In BMI task 1 (jumper task), animals needed to generate neural activity that would be decoded as locations they wanted to move to so that they could reach each goal (to obtain a reward). In BMI task 2 (Jedi task), animals were fixed at the center of the virtual arena (but could rotate) and needed to generate activity corresponding to locations where they wanted an external object to move to so that the object reached the goal, then they needed to sustain that activity to maintain the object there (to maximize reward). (**B**) Schematic of the VR system (left). The animal was free to rotate its body in the horizontal plane. In the running task, the animal's location in the virtual arena environment was updated according to treadmill movement. Simultaneously recorded spiking from a population of hippocampal CA1 units expressed place fields—the basis of the cognitive map of the environment (right). Decoder was then trained using binned spiking activity and location data. (**C**) In both BMI tasks, the treadmill no longer updated VR. Instead, the animal or object location was controlled solely by real-time hippocampal activity. A neural signal processor rapidly assigned activity to individual units, whose spike counts were fed into the decoder. VR projection was updated according to locations output by the decoder. In the jumper (Jedi) task, the animal's (object's) virtual location was moved toward the most recent decoded locations. PC, personal computer.

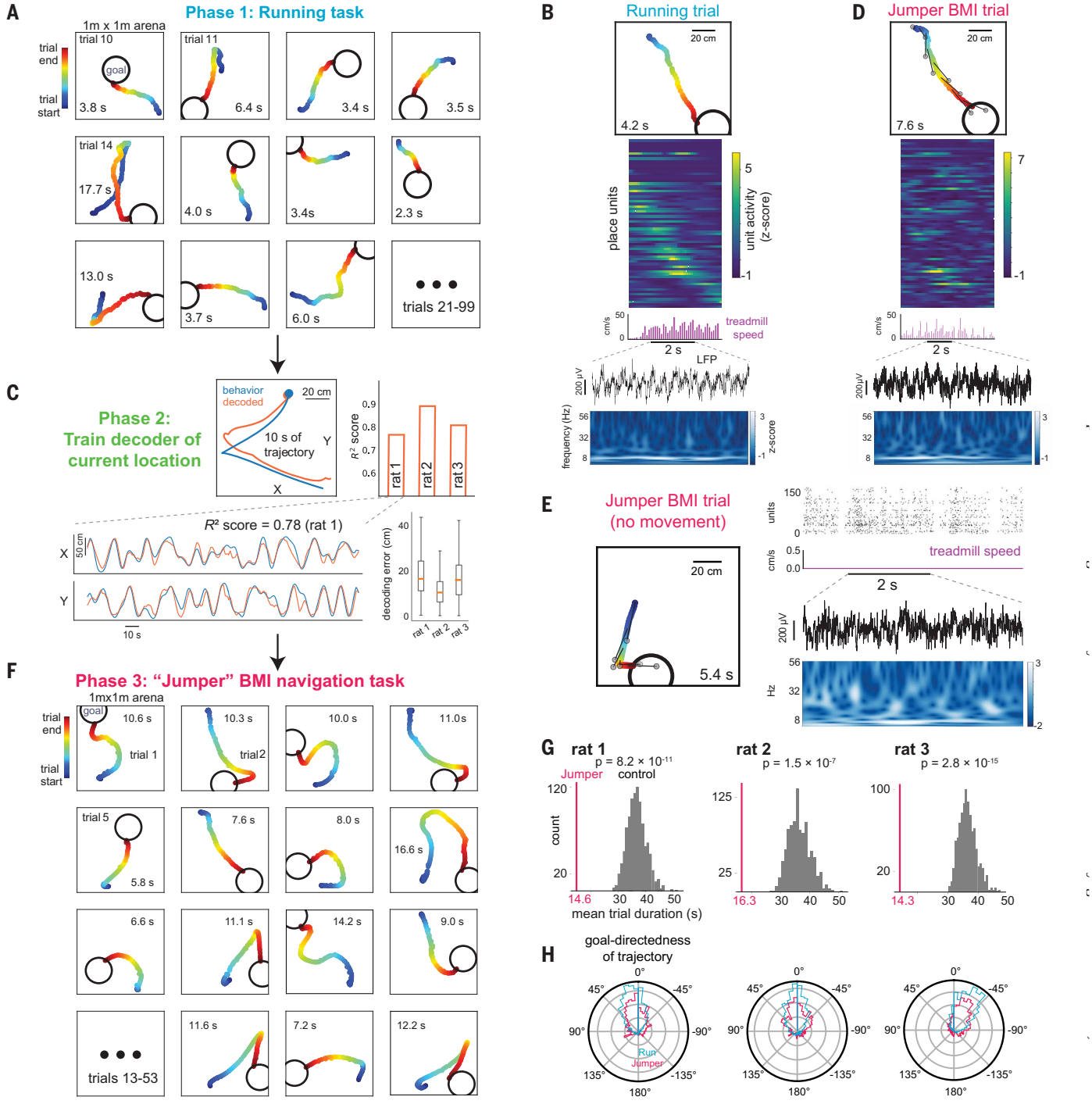


Fig. 2. Rats can navigate to goals by controlling their hippocampal activity. In both the running and jumper BMI tasks, animals were rewarded when they reached each goal. (**A**) Animal trajectories in the virtual arena for consecutive running task trials. Trial duration (time to reach goal) in seconds is shown. (**B**) Example running task trial. From top: trajectory, firing rate (z-scored) of individual units (units were ordered by time of peak activity), treadmill speed, and LFP from one recording channel and corresponding wavelet spectrogram during trial. (**C**) Accuracy of trained decoder of animal's current location for held-out running task data. Actual and decoded trajectories during example trial (top left) and across several trials (for *x* and *y* coordinates separately, bottom left). Median decoding error (distance between actual and decoded locations) with range and quartiles (bottom right). (**D**) Example jumper BMI trial with similar trajectory as the running trial in (B). From top: trajectory generated by the animal controlling its hippocampal activity and the decoder output (animal is teleported toward decoded location; each gray circle represents the decoded location at the time the animal is at the corresponding point in the trajectory connected by the dark line, sampled here every 1 s), firing rate of individual units [using same order of units as in (B)], treadmill speed, LFP, and spectrogram. (**E**) Example jumper BMI trial in which animal did not move the treadmill. Trajectory (left) as in (D). Right, from top: unit activity, treadmill speed, LFP, and spectrogram. See fig. S10 for all 10 nonmovement trials. (**F**) BMI-generated trajectories for consecutive jumper trials. (**G**) Mean jumper trial duration (magenta vertical line) is significantly lower than distribution of expected mean duration for simulated trials if goals were in random locations. (**H**) Polar distribution of angle between direction of movement and direction to goal during running and jumper tasks. Zero corresponds to animal movement directly toward the goal center.

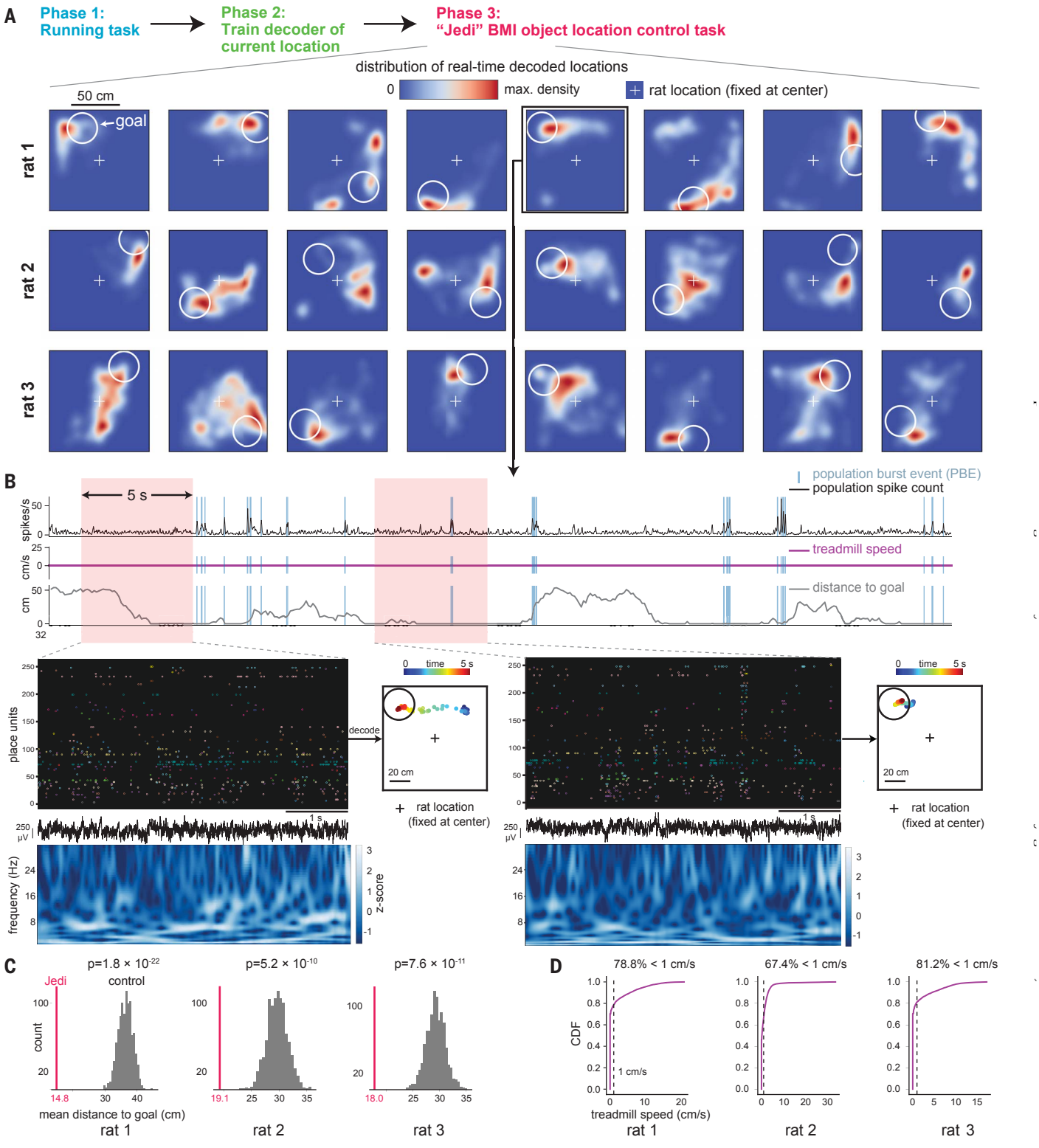


Fig. 3. Rats can move objects to remote goal locations and maintain them there by controlling their hippocampal activity. In the Jedi BMI task, trials did not end when the external controlled object first reached the goal; instead, animals were rewarded as long as the object was in the goal region (white circle), for up to 3 min per trial. The animal was always fixed at center of the virtual arena but could rotate its body and generally turned toward each goal. (**A**) Distribution of real-time decoded locations (output every 100 ms) generated by the animal controlling its hippocampal activity across eight consecutive Jedi BMI trials for rats 1, 2, and 3. Panels show decoded locations during each trial (up to 3 min; fig. S11). Periods when the animal's body rotated >12°/s were excluded. See text and methods for details. The external controlled object (which was visible for rats 1 and 2 but invisible for rat 3) was moved toward the decoded location (fig. S11 shows that the distribution of object locations was essentially the same as the distribution of decoded

locations). (**B**) A 40-s-long period during an example trial during which the animal did not move the treadmill. From top: summed activity across all units with PBEs identified, treadmill speed, distance of decoded location from goal (0 means inside goal region), and close-ups of two 5-s periods [as animal moves object to goal (left) and as animal maintains object at goal (right); points in the arena represent sequence of decoded locations] with spike trains of units, LFP, and spectrogram. See fig. S12 for additional example periods. (**C**) Mean distance of decoded location from goal across all trials (magenta vertical line) is significantly lower than mean distance expected for randomized goal locations. (**D**) Treadmill speed distribution during periods shown in (A), illustrating that the animal was generally still during task performance.

ended and a new goal cue appeared at a random location.

Rats successfully navigated by controlling their hippocampus, generating efficient paths to each goal (Fig. 2F; see figs. S3 to S5 for all trials of three rats, and figs. S6 to S8 for all trials re-decoded using a shorter decoding window and without smoothing). To check whether this performance could be attributed to non-spatially-specific neural activity (e.g., modulating global firing rate), we randomly shuffled the spike trains across place units, ran the shuffled data through the original decoder to produce simulated trajectories, then determined how long it would have taken to reach the same sequence of goal locations as in the original experiment. Shuffled-unit mean trial durations were much longer than the actual means ($P < 10^{-100}$, three rats, one session each), suggesting that performance depended on generating place field–related activity. To test whether generating non-goal-directed sequences of location-specific activity (e.g., random movement within the cognitive map) could explain the performance, we randomly shuffled the goal locations in each trial while preserving the original BMI trajectories and then determined the time that would have been needed to reach the shuffled goals. Shuffled-goal mean trial durations were again much longer than actual means ($P = 2.8 \times 10^{-15}$ to 1.5×10^{-7}) (Fig. 2G), indicating that animals' BMI trajectories were clearly goal-directed. Goal-directedness was also apparent from the distribution of angles between the animal's instantaneous direction of BMI-generated movement and the direction from the animal's current location to the goal, which was concentrated around a value near 0° (Fig. 2H). Thus, even though jumper trials took longer than running trials (mean trial duration across animals: 15.1 s versus 6.9 s; note, however, that BMI decoding and smoothing added a few seconds to jumper durations), the animals' routes revealed effective, goal-directed, map-based BMI navigation. Furthermore, such performance was achieved without extensive BMI training (Fig. 2F; figs. S3 to S5 show sessions 3, 9, and 2 for rats 1, 2, and 3, respectively; table S1; a fourth rat failed to perform either BMI task).

Although animals were free to physically run during the jumper task, such movement was not necessary for task performance. Initially, animals ran as in the running task, but in later trials, animals ran less (fig. S9). In a subset of trials (10 out of 161 trials) (Fig. 2E and fig. S10), animals remained still, yet in all cases they efficiently reached the goal. Moreover, this successful navigation did not depend on activity in population burst events (PBEs), which often appear during immobility and during which brief activation of place cell representations for remote locations has been shown to occur (*13*–*16*, *18*, *20*, *21*, *23*).

BMI object location control task

Although episodic memories are encoded and often retrieved using a first-person perspective, individuals can also imagine scenarios from a third-person perspective, with other animate and inanimate players taking part. Furthermore, imagination often involves holding a single thought in mind for extended periods. Therefore, our second BMI task, the Jedi task, tested whether animals—while remaining in the same place—could use the same map of the arena to control the location of a virtual object, guide it to the goal cue location, and maintain it nearby. The jumper and Jedi tasks thus used different forms of feedback: self-location and the location of an object, respectively. After the same running task and decoder training phases as in the jumper experiment, the animals in Jedi were fixed (but could freely turn) at the arena's center, and the object's location was updated to the BMI-decoded location (with a 2-s smoothing window). In each trial, the goal cue remained in the same place, providing reward as long as the object touched it. After 3 min or the rat having received 0.5 ml of reward in total, whichever came first, a new goal cue appeared at a distant random location for the next trial.

Rats could activate and sustain a remote location's representation around the goal for long periods, until the trial ended, and then shift attention to the next goal (Fig. 3, A and B; fig. S11; and movie S1). Performance was measured using the mean distance (over time) between the decoded locations and goals. Shuffling spike trains across units yielded much greater mean distances than the actual means ($P = 2.2 \times 10^{-5}$ to 2.6×10^{-3}, three rats, one session each). To assess the goal-directedness of BMI-generated activity, we shuffled the goal locations while preserving the locations output by the decoder. The decoded (and controlled object's) location was far more concentrated around the actual remote goal cue than shuffled goal locations ($P = 1.8 \times 10^{-22}$ to 5.2×10^{-10}) (Fig. 3C and fig. S11), indicating clear goal-directed control of activity. Again, such performance occurred without extensive training (Fig. 3A shows sessions 7, 6, and 3 for animals 1, 2, and 3, respectively). Task performance was not dependent on PBEs, as there was no change in performance when all activity in PBEs was eliminated and the decoder was rerun post hoc (fig. S11).

Animal movement was generally low when engaged in the Jedi task (Fig. 3D), and movement was not required for successful performance. There were many longer periods (≥8 s long with a treadmill speed of ≤1 cm/s, 38 periods, mean: 17.3 s, maximum: 44.0 s) during which the animal did not move the treadmill while it directed the object to the goal and/or held it there (34 of 38 periods) (Fig. 3B and fig. S12). Activity during PBEs was also generally not necessary for performance in these non-movement segments (fig. S12).

Features of volitionally generated spiking and local field potential activity

What characteristics did the volitionally generated activity have? First, mean firing rates per unit were similar between jumper and running tasks (fig. S13A). Mean firing rates per unit were correlated across Jedi and running tasks, but lower in Jedi (fig. S13B)—consistent with the decreased physical movement in Jedi.

We then investigated the hypothesis that, to move themselves or the object toward a given (decoded) location in the jumper and Jedi tasks, animals generated a pattern of firing rates across units (i.e., a population vector, or PV) similar to the mean PV at that location over the entire running task (called the reference PV, or rPV) (Fig. 4A). (Note that the set of rPVs for all locations is thus equivalent to the standard place field map across the population.) We examined the correlation between the PV generated at each moment (in every 500-ms window) during jumper or Jedi and the rPV of the decoded location at that moment. As a benchmark, we computed the correlation between the 500-ms PVs during the running task with the rPVs corresponding to the animal's actual location at those times (Fig. 4, B and C, "Run") as well as the correlation between the running task PVs and the rPVs of random locations (Fig. 4, B and C, "randRun"). We then correlated jumper or Jedi PVs with the rPVs of the decoded locations at each moment (Fig. 4B for "jumper" and Fig. 4C and fig. S14 for "Jedi") and with rPVs of random locations ("randJumper" and "randJedi"). Jumper and Jedi PVs were significantly correlated

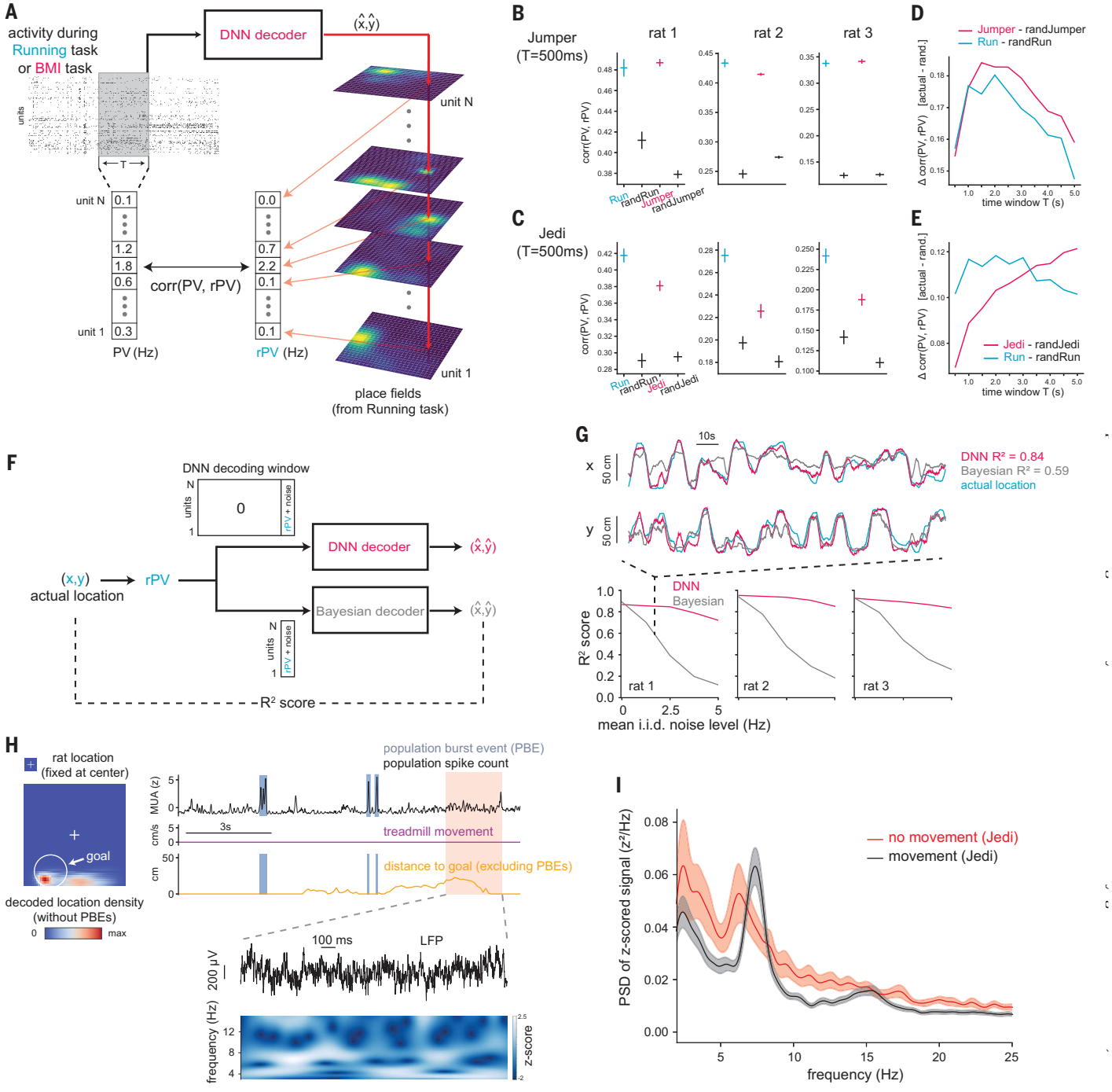


Fig. 4. Volitionally generated nonlocal activity is similar to the activity when the animal is at the corresponding locations and is associated with theta-band power in the LFP. (**A** to **E**) The population vector (PV) of ongoing spiking activity was compared with the average place field activity (rPV) at a given location during the running task. (A) Schematic of comparison. (B) Mean correlation of instantaneous (500-ms window) PV during running or jumper task with rPV for the current location (in the running task), current decoded location (in the jumper task), or random location in the running (randRun) or jumper (randJumper) task. (C) Same as (B) but for the Jedi task. For Jedi, only periods when decoded location was near (within 5 cm of) the goal were included [also for (E)]. [(D) and (E)] Correlation of PV with rPV relative to baseline random value as a function of the time integration window for determining the PV. (**F** and **G**) Evaluation of decoder performance when ground truth activity for each location, i.e., the rPV, was input into the decoder. (F) Schematic of evaluation procedure. (G) Comparison of our DNN decoder to Bayesian decoder for different levels of added noise, with example traces using a specific level of noise (top). (**H**) Distribution of decoded location (left) during Jedi task segment with no treadmill movement (right). Right, from top: summed activity across all units with PBEs identified, treadmill speed, distance of decoded location (excluding data during PBEs) from goal, and close-up of LFP with spectrogram. (**I**) Power spectral density of z-scored (for pooling across animals) LFP during the Jedi task for periods of treadmill movement and all long segments (≥8 s) without treadmill movement. See text and methods in the supplementary materials for details. Here and elsewhere, all confidence intervals (CIs) are 95% CIs.

with the rPVs associated with the decoded locations versus random locations, consistent with the hypothesis. Furthermore, jumper PV-rPV correlations were comparable to running task PV-rPV correlations. In line with this, the example running (Fig. 2B) and jumper (Fig. 2D) trials, which happened to share similar trajectories, showed similar activity patterns across place units over time. PV-rPV correlation scores were, unlike in jumper, lower in Jedi than in the running task using 500-ms windows (Fig. 4C), consistent with noisier generation of nonlocal representations and/or lower firing rates (fig. S13B) in Jedi. However, with longer integration windows (>500 ms) (Fig. 4, D and E), the PVs generated during Jedi matched the rPVs as well as the best match during the running task (note that longer integration times work for Jedi because animals activated goal location representations for extended periods). These results indicate that, during BMI task performance, animals generated nonlocal population activity as similar to the corresponding place field representations as when they actually visited those locations in the running task. Were these place field–like (i.e., rPV-like) patterns what our deep network detected to decode location? While determining what features a deep network uses for decoding is generally not straightforward, inputting a single location's rPV for a brief duration was sufficient to produce accurate location decoding (Fig. 4, F and G), consistent with the decoder being tuned to detect rPV-like activity. In addition, unlike the commonly used Bayesian decoder (*45*), our decoder was highly robust to noise (Fig. 4G) by design because of the use of data augmentation during training.

Lastly, we analyzed the local field potential (LFP) activity during BMI task performance (Fig. 4, H and I). When animals move, the rodent hippocampal LFP is known to display prominent theta band (~5 to 12 Hz) power, which peaked at ~7.3 Hz during periods of movement in the running and BMI tasks (Fig. 4I). During the extended periods of nonmovement when the animal was performing the Jedi task, the theta peak shifted down to 6.3 Hz (Fig. 4I). Note that, unlike the more continuous theta oscillations during movement, the oscillations during such nonmovement periods tended to be more intermittent.

Discussion

Previous BMI research has yielded major advances in the control of robotic arms, computer cursors, and other devices by activity from the primary motor cortex, premotor cortex, and posterior parietal cortex (*24*–*35*). The hippocampal cognitive map has a code that represents space in terms of absolute location in the external environment versus location relative to (e.g., in front of, or to the right or left of) the animal (*8*–*11*), and it was unknown whether a subject could control a BMI by means of this code. With this study, we demonstrated a hippocampal map–based BMI in which the subject is able to control its location or that of other objects by activating location representations in terms of absolute space, independent of where the animal currently is. That is, even though animals generally (but not always) turned their body toward the goal, the activity that needed to be generated differed depending on the location of the goal with respect to the environment. The relatively small amount of training needed for the animals to perform our BMI tasks is in line with our use of a biomimetic decoder (*35*, *46*), that is, one based on the neural code that the subject naturally employs.

In humans, imagining or recalling objects or video clips is accompanied by hippocampal activity in individual neurons similar to that when viewing the original stimuli (*47*, *48*). This suggests that the mechanisms allowing animals to selectively activate their nonlocal hippocampal spatial representations, as we have shown here, could also underlie our ability to actively recall or imagine experiences in other places. The ability of rodents to perform these BMI tasks should thus allow imagination, as well as the voluntary recall of memory, to be investigated using the range of tools available for this model system. More generally, the neural processes engaged here could underlie our capacity to perform "mental time travel"—travel back in time by reexperiencing richly detailed episodic memories and travel forward in time by generating possible future scenarios (*49*). Mental time travel depends critically on the hippocampus (*4*–*6*, *50*–*52*) and enables subjects to internally simulate new experiences according to their world model. This can aid decision-making and facilitate learning in complex situations where trial and error is expensive, as shown using artificial agents (*3*, *53*–*55*).

Along these lines, the rats in our study could control their hippocampal map-based activity on a timescale of seconds, corresponding to the speed and duration at which humans relive past events or imagine new scenarios. Navigational trajectories each lasted ~10 s, and a virtual object could be held at a remote location for several seconds. This contrasts with the previously described fast (~100 ms) sequences of nonlocal hippocampal activity in awake rodents (i.e., awake replay events, which are associated with population bursts and sharp wave–ripples) thought to be associated with planning (*12*, *16*, *18*, *21*), and which were not responsible for the performance in our BMI tasks (analysis in which all PBEs were removed). The content of such replay events, which can portray specific routes through the environment starting from the animal's current location, has been shown to be correlated with deliberative (*12*) and future (*16*, *18*, *21*) behavior. However, it is not known whether this content is—or replay content in general can be—under an animal's volitional control. For instance, hippocampal activity displays similar fast sequences during sleep (*23*), thus nonlocal path generation per se does not appear to require intention. If awake replay is volitionally controlled, these events could represent a brief consideration of alternatives for making a quick decision and be distinct from the more comprehensive mental simulations of possible scenarios that take seconds. Previous work has also described neurons in the hippocampus and related areas whose activity is tuned to the angle to a goal or a salient cue or object relative to the direction the animal is facing (*56*–*59*). In addition, hippocampal neurons that are tuned to the location of conspecifics have been found (*60*, *61*). As with fast sequences, whether these forms of activity that reflect locations away from the animal are volitionally controlled is yet to be determined.

Beyond aiding decision-making, the ability to control the content of the hippocampal spatial and episodic memory system could help explain the richness of our inner lives. Finally, the ability to control hippocampal activity to guide oneself or objects to intended locations—and do so with high signal-to-noise readout using our decoder—could lead to new BMI applications for restoring or enhancing function by realizing a subject's high-level intentions with respect to their internal world models.

REFERENCES AND NOTES

1. D. L. Schacter, D. R. Addis, R. L. Buckner, *Nat. Rev. Neurosci.* **8**, 657–661 (2007).
2. D. L. Schacter *et al*., *Neuron* **76**, 677–694 (2012).
3. J. B. Hamrick, *Curr. Opin. Behav. Sci.* **29**, 8–16 (2019).
4. D. Hassabis, D. Kumaran, S. D. Vann, E. A. Maguire, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1726–1731 (2007).
5. D. Hassabis, D. Kumaran, E. A. Maguire, *J. Neurosci.* **27**, 14365–14374 (2007).
6. D. R. Addis, A. T. Wong, D. L. Schacter, *Neuropsychologia* **45**, 1363–1377 (2007).
7. E. C. Tolman, *Psychol. Rev.* **55**, 189–208 (1948).
8. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Clarendon Press, 1978).
9. J. O'Keefe, J. Dostrovsky, *Brain Res.* **34**, 171–175 (1971).
10. J. O'Keefe, *Prog. Neurobiol.* **13**, 419–439 (1979).
11. M. A. Wilson, B. L. McNaughton, *Science* **261**, 1055–1058 (1993).
12. A. Johnson, A. D. Redish, *J. Neurosci.* **27**, 12176–12189 (2007).
13. K. Diba, G. Buzsáki, *Nat. Neurosci.* **10**, 1241–1242 (2007).
14. T. J. Davidson, F. Kloosterman, M. A. Wilson, *Neuron* **63**, 497–507 (2009).
15. M. P. Karlsson, L. M. Frank, *Nat. Neurosci.* **12**, 913–918 (2009).
16. B. E. Pfeiffer, D. J. Foster, *Nature* **497**, 74–79 (2013).
17. A. M. Wikenheiser, A. D. Redish, *Nat. Neurosci.* **18**, 289–294 (2015).
18. C.-T. Wu, D. Haggerty, C. Kemere, D. Ji, *Nat. Neurosci.* **20**, 571–580 (2017).
19. K. Kay *et al*., *Cell* **180**, 552–567.e25 (2020).
20. X. Mou, A. Pokhrel, P. Suresh, D. Ji, *Neuron* **110**, 891–902.e7 (2022).
21. J. Widloski, D. J. Foster, *Neuron* **110**, 1547–1558.e8 (2022).
22. K. Louie, M. A. Wilson, *Neuron* **29**, 145–156 (2001).
23. A. K. Lee, M. A. Wilson, *Neuron* **36**, 1183–1194 (2002).
24. J. Wessberg *et al*., *Nature* **408**, 361–365 (2000).
25. J. M. Carmena *et al*., *PLOS Biol.* **1**, E42 (2003).
26. S. Musallam, B. D. Corneil, B. Greger, H. Scherberger, R. A. Andersen, *Science* **305**, 258–262 (2004).
27. L. R. Hochberg *et al*., *Nature* **442**, 164–171 (2006).

28. G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, K. V. Shenoy, *Nature* **442**, 195–198 (2006).
29. M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, A. B. Schwartz, *Nature* **453**, 1098–1101 (2008).
30. P. T. Sadtler *et al*., *Nature* **512**, 423–426 (2014).
31. T. Aflalo *et al*., *Science* **348**, 906–910 (2015).
32. G. K. Anumanchipalli, J. Chartier, E. F. Chang, *Nature* **568**, 493–498 (2019).
33. F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, *Nature* **593**, 249–254 (2021).
34. E. E. Fetz, *J. Physiol.* **579**, 571–579 (2007).
35. M. A. Lebedev, M. A. L. Nicolelis, *Physiol. Rev.* **97**, 767–837 (2017).
36. D. Ishikawa, N. Matsumoto, T. Sakaguchi, N. Matsuki, Y. Ikegaya, *J. Neurosci.* **34**, 5044–5053 (2014).
37. K. Patel, C. N. Katz, S. K. Kalia, M. R. Popovic, T. A. Valiante, *Brain* **144**, 3651–3663 (2021).
38. M. Cerf *et al*., *Nature* **467**, 1104–1108 (2010).
39. D. Aronov, D. W. Tank, *Neuron* **84**, 442–456 (2014).
40. C. Lai, thesis, University of Cambridge (2020).
41. C. Lai *et al*., bioRxiv 2023.09.14.557854 [Preprint] (2023); https://doi.org/10.1101/2023.09.14.557854.
42. C. D. Harvey, F. Collman, D. A. Dombeck, D. W. Tank, *Nature* **461**, 941–946 (2009).
43. G. Chen, J. A. King, Y. Lu, F. Cacucci, N. Burgess, *eLife* **7**, e34789 (2018).
44. J. S. Lee, J. J. Briguglio, J. D. Cohen, S. Romani, A. K. Lee, *Cell* **183**, 620–635.e22 (2020).
45. K. Zhang, I. Ginzburg, B. L. McNaughton, T. J. Sejnowski, *J. Neurophysiol.* **79**, 1017–1044 (1998).
46. K. V. Shenoy, J. M. Carmena, *Neuron* **84**, 665–680 (2014).
47. G. Kreiman, C. Koch, I. Fried, *Nature* **408**, 357–361 (2000).
48. H. Gelbard-Sagiv, R. Mukamel, M. Harel, R. Malach, I. Fried, *Science* **322**, 96–101 (2008).
49. E. Tulving, *Psychologie canadienne* **26**, 1–12 (1985).
50. F. Andelman, D. Hoofien, I. Goldberg, O. Aizenstein, M. Y. Neufeld, *Neurocase* **16**, 426–435 (2010).
51. E. Race, M. M. Keane, M. Verfaellie, *J. Neurosci.* **31**, 10262–10269 (2011).
52. T. Bartsch, J. Döhring, A. Rohr, O. Jansen, G. Deuschl, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17562–17567 (2011).
53. T. Weber *et al*., *Adv. Neural Inf. Process. Syst.* **30**, 5690–5701 (2017).
54. D. Ha, J. Schmidhuber, arXiv:1803.10122 [cs.LG] (2018).
55. D. Hafner, T. Lillicrap, J. Ba, M. Norouzi, arXiv:1912.01603 [cs.LG] (2019).
56. L. Acharya, Z. M. Aghajan, C. Vuong, J. J. Moore, M. R. Mehta, *Cell* **164**, 197–207 (2016).
57. A. Sarel, A. Finkelstein, L. Las, N. Ulanovsky, *Science* **355**, 176–180 (2017).
58. C. Wang *et al*., *Science* **362**, 945–949 (2018).
59. J. Ormond, J. O'Keefe, *Nature* **607**, 741–746 (2022).
60. T. Danjo, T. Toyoizumi, S. Fujisawa, *Science* **359**, 213–218 (2018).
61. D. B. Omer, S. R. Maimon, L. Las, N. Ulanovsky, *Science* **359**, 218–224 (2018).
62. C. Lai, S. Tanaka, T. D. Harris, A. K. Lee, Volitional activation of remote place representations with a hippocampal brain-machine interface, version 1.0, Zenodo (2023); https://doi.org/10.5281/zenodo.8360872.

ACKNOWLEDGMENTS

We thank M. Bolstad for work on the virtual reality software; W. Bishop, S. Romani, and X. Zhao for valuable discussions regarding the study; S. Lindo, R. Gattoni, and other members of the Janelia Vivarium team, and B. Lustig, A. Sohn, S. Sawtelle, I. Negrashov, B. Foster, J. Osborne, J. Arnold, R. Rogers, and J. Cohen for technical assistance; and J. Dudman, S. Romani, A. Hantman, T. Wang, and J. Colonell for valuable comments on the manuscript. **Funding:** This work was supported by the Howard Hughes Medical Institute (A.K.L. and T.D.H.).
Author contributions: Conceptualization: C.L. and A.K.L. Task design: C.L. BMI and task software and BMI hardware: C.L. Behavioral and recording hardware: S.T. Experiments: S.T. and C.L. Data analysis: C.L. and A.K.L. Writing: C.L., A.K.L., and S.T. Supervision: A.K.L. and T.D.H. **Competing interests:** T.D.H. is also affiliated with the Department of Biomedical Engineering at Johns Hopkins University. **Data and materials availability:** All data needed to assess the findings in this study are publicly available at Zenodo (*62*). All analysis methods are described in the main text and supplementary materials, and the code for building and applying the decoder is publicly available at Zenodo (*62*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh5206
Materials and Methods
Figs. S1 to S14
Table S1
References (*63*–*72*)
MDAR Reproducibility Checklist
Movie S1

Submitted 9 March 2023; accepted 22 September 2023
10.1126/science.adh5206

ELECTROCHEMISTRY

Design of a trigonal halide superionic conductor by regulating cation order-disorder

Seungju Yu[1]†, Joohyeon Noh[1]†, Byunghoon Kim[1], Jun-Hyuk Song[1]‡, Kyungbae Oh[1], Jaekyun Yoo[1], Sunyoung Lee[1], Sung-O Park[1], Wonju Kim[1], Byungwook Kang[1], Donghyun Kil[1,2], Kisuk Kang[1,3,4,5]*

Lithium-metal-halides have emerged as a class of solid electrolytes that can deliver superionic conductivity comparable to that of state-of-the-art sulfide electrolytes, as well as electrochemical stability that is suitable for high-voltage (>4 volt) operations. We show that the superionic conduction in a trigonal halide, such as Li_3MCl_6 [where metal (M) is Y or Er], is governed by the in-plane lithium percolation paths and stacking interlayer distance. These two factors are inversely correlated with each other by the partial occupancy of M, serving as both a diffusion inhibitor and pillar for maintaining interlayer distance. These findings suggest that a critical range or ordering of M exists in trigonal halides, and we showcase the achievement of high ionic conductivity by adjusting the simple M ratio (per Cl or Li). We provide general design criteria for superionic trigonal halide electrolytes.

Lithium-metal-halides have recently been revisited as promising solid electrolyte (SE) candidates that can provide solutions to the interfacial instability issues found in sulfide- and oxide-based SEs (*1*–*3*). Most halide-based SEs commonly exhibit high oxidation potential and a wide electrochemical stability window (*3*–*5*), which help mitigate instability problems in sulfide-based SEs (*6*). Furthermore, their exceptional ductility, comparable to that of sulfide-based SEs, provides process advantages in ensuring an intimate interface with the electrode (*3*, *7*), which contributes to resolving instability issues in oxide-based SEs (*8*). Although early studies on halide-based SEs indicated that poor ionic conductivities (typically <~10^{-6} S cm^{-1}) constituted a major bottleneck (*3*, *9*, *10*), recent breakthroughs have revealed that the ionic conductivity can be boosted by identifying appropriate synthetic routes, which introduces a wide range of halide-based Li_3MX_6 (where M is Y, Er, or In, and X is Cl, Br, or I) (*3*, *7*, *11*). Notably, moving from the classical ampule synthesis to mechanochemical synthesis substantially improved room-temperature ionic conductivity up to ~10^3 orders of magnitude for the Li_3MCl_6 (where M is Y or Er) (*7*, *12*–*15*). It was speculated that the higher degree of cation disordering from the mechanochemical synthesis could open up the bottleneck transition triangular area for lithium diffusion (*12*, *15*). Although the factors leading to the hidden superionic conductivity of halide SEs remain unclear, these previous findings strongly suggest a close interplay between superionic conduction and cation arrangements and occupancies that are attainable from various synthetic methods. Hence, understanding the fundamental relationship between the superionic conduction mechanism and structural factors in trigonal Li_3MCl_6 halides is critical for realizing important advancements in their ionic conductivities. In this work, we used first-principles calculations and experiments to investigate the conduction mechanism and propose and validate our design strategies for enhancing ionic conductivity using Li_3YCl_6 (LYC) as a model system.

Diverse cation arrangements and corresponding lithium-diffusion kinetics

LYC adopts a layered-type structure with a hexagonal close-packed (hcp) anion framework in the $P\bar{3}m1$ space group (Fig. 1A) (*15*). The Y ion occupies octahedral sites at the 1a and 2d Wyckoff positions, referred to as Y 1a and Y 2d, respectively. Previous studies (*12*, *15*) show that the partial occupancy of Y 1a is always close to 1, whereas that of Y 2d depends on synthetic routes, ranging from 0.03 to 0.97. We refer to the layer containing Y 1a as layer 1, and layer 2 is located between the layers of layer 1 (left panel of Fig. 1A). High-valent Y is surrounded by six Li or vacancy octahedra, forming honeycomb-type in-plane ordering (right panel of Fig. 1A). Distinct partial occupancies of Y result in two cationic column types along the *c* axis: α columns with alternating Y 1a and vacant sites, and β columns with repeated 2d sites (middle panel of Fig. 1A). Various cation arrangements are possible within the β column, from complete randomization of Y 2d site occupations to an ordered arrangement equivalent to that of the α column. Nevertheless, two neighboring Y 2d sites cannot be simultaneously occupied because of a strong face-sharing electrostatic interaction, which limits the number of local β configurations (supplementary text, section 1).

To understand the relationship between cation arrangement and ionic conductivity, we first considered an ideal ααα-model structure with a β configuration that is identical to the α-type arrangement in partial occupancy (Fig. 1B). This structure features alternating Y-free lithium layers and Y-rich layers, referred to as Y-free ordering and Y-3 ordering, respectively. Additionally, we considered a pseudo-disordered αββ-model structure with 2d sites occupied in layer 2 and unoccupied in layer 1 (Fig. 1C), which produces two distinct in-plane cation configurations: triangular ordering between two Y ions and one vacancy (Y-2 ordering) and ordering among one Y ion and two vacancies (Y-1 ordering). Importantly, our two models include all possible local in-plane Y orderings surrounding lithium ions (blue triangles in Fig. 1, B and C) within the $P\bar{3}m1$ crystal symmetry, and their combinations can represent any global structure of stoichiometric Li_3YCl_6 that bears partial occupancies of Y in 2d sites (supplementary text, section 2).

The lithium-diffusion kinetics of the two models were comparatively examined using ab initio molecular dynamic (AIMD) calculations. Figure 1D illustrates the room-temperature ionic conductivity, extrapolated to 300 K from AIMD calculations at 40-K intervals between 500 and 700 K (see methods and fig. S2). Both ααα and αββ models exhibit anisotropic diffusion, with *c*-axis conduction that is much faster than *ab*-plane conduction, consistent with previous reports (*4*, *16*). Despite fast lithium conduction along the *c* direction, the total ionic conductivities remain lower for both cases, indicating that the *ab*-plane component limits overall conductivity. Interestingly, the αββ-model structure with higher *ab*-plane conductivity (6.83 versus 0.70 mS cm^{-1}) and lower *c*-axis conductivity (102 versus 114 mS cm^{-1}) showed higher overall conductivity (26.3 versus 18 mS cm^{-1}), underscoring the importance of *ab*-plane ionic conductivity as the rate-limiting factor.

Figure 1E visualizes the anisotropic lithium diffusion networks in LYC using isosurface plots of Li probability density from AIMD calculations at 500 K. Both the ααα and αββ models show well-connected *c*-axis lithium diffusion networks, which enables high ionic conductivity through continuously interconnected octahedral site hopping. By contrast, the *ab*-plane lithium networks differ in the two model systems. Lithium

[1]Department of Materials Science and Engineering, Research Institute of Advanced Materials (RIAM), Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. [2]SDI R&D Center, Samsung SDI Co., Ltd., Yeongtong-gu, Suwon-si, Gyeonggi-do 16678, Republic of Korea. [3]Center for Nanoparticle Research, Institute for Basic Science (IBS), Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. [4]Institute of Engineering Research, College of Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. [5]School of Chemical and Biological Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea.
*Corresponding author. Email: matlgen1@snu.ac.kr
†These authors contributed equally to this work.
‡Present address: LiB Materials Research Group, Research Institute of Industrial Science and Technology (RIST), 100 Songdogwahak-ro, Yeonsu-gu, Incheon, Republic of Korea.

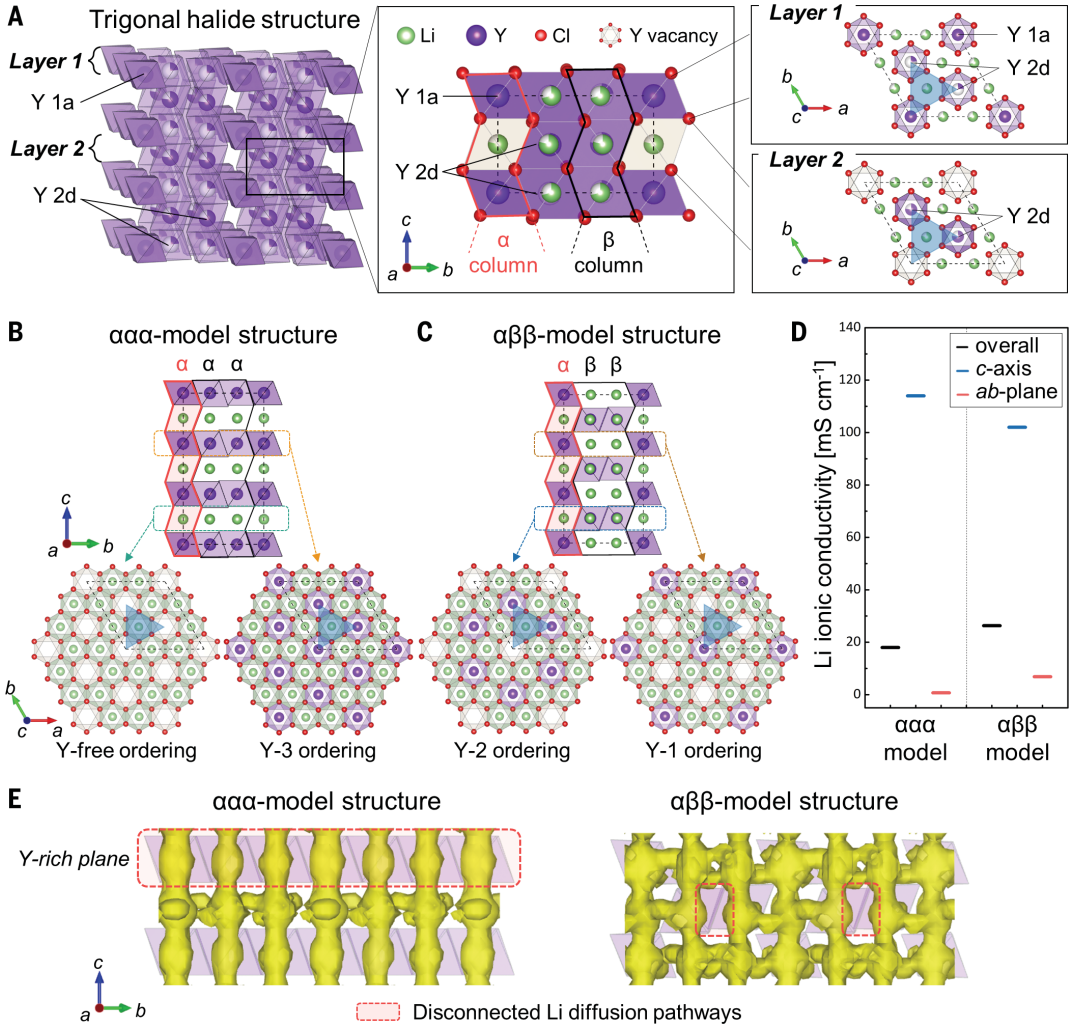


Fig. 1. Comparison of ionic conductivities and structures of two representative models that have different cation arrangements of Li_3YCl_6. (**A**) Structure of Li_3YCl_6 (*15*), which has Y-ion partial occupancy on the Wyckoff 2d site. The overall connectivity and arrangements of the Y 1a and Y 2d sites are shown in the panel on the left. The side view of Li_3YCl_6 unit cell is illustrated in the middle panel. The panel on the right shows the configuration of each plane (i.e., layers 1 and 2). (**B** and **C**) Two different cation arrangement models for a 1-by-1-by-3 supercell of Li_3YCl_6, which represent the models in which the Y ions in the Wyckoff 1a and 2d sites are in the same (B) and different (C) layers. Each model further consists of two different *ab*-plane orderings: Y-free and Y-3 orderings for the ααα-model structure (B) and Y-1 and Y-2 orderings for the αββ-model structure (C). The Y-free, Y-1, Y-2, and Y-3 orderings contain 0, 1, 2, and 3 Y ions per unit cell, respectively, as indicated by the blue triangles and also shown in the right panel of (A). (**D**) Each component of extrapolated ionic conductivity at 300 K from AIMD simulations for the ααα- and αββ-model structures (black for overall, blue for *c* axis, and red for *ab* plane). (**E**) Isosurface (yellow) of lithium probability density $P = P_{max}/256$ at 500 K of two model structures to present the lithium-diffusion pathway. The red dashed rectangle indicates the disconnected lithium-diffusion pathway along the *ab* plane.

ions are connected only in every other plane in the ααα model, with interchain connections occurring exclusively through the Y-free plane, as shown by the red dashed box for the disconnected region. However, lithium conduction is feasible in all *ab* planes of the αββ model, offering interchain connections throughout. These selective diffusion paths are speculated to contribute to the different conduction properties of the two models. Nevertheless, this does not fully account for the nearly order-of-magnitude higher *ab*-plane lithium conductivity in the αββ model compared with the ααα model. It is particularly perplexing because the Y-free layer still offers interchain connection in the ααα model, and lithium diffusion has been presumed to be easier in a lithium layer that is free of transition metals, as has been extensively studied in layered transition-metal oxides such as $LiNiO_2$ (*17*, *18*).

Percolating the in-plane diffusion network and Y ordering

We further investigated lithium diffusion in the *ab* plane, considering various local cation (Y) configurations surrounding lithium ions. Figure 2A presents in-plane views of Li probability density plots, obtained from AIMD calculation at 500 K, for each Y ordering. These plots illustrate that lithium ions can freely roam within the *ab* plane for Y-free and Y-1 orderings, whereas Y-2 and Y-3 orderings feature disconnected intralayer networks, resulting in isolated local lithium-diffusion paths (red dashed lines). Interestingly, lithium-ion immobility in the Y-3 layer is observed with entirely isolated intralayer diffusion paths, which signifies that lithium ions can only migrate through the *c*-axis diffusion channel, making the layer itself the bottleneck of diffusion. Similarly, the percolation diffusion network in the Y-2 layer is restricted within a small circular ring

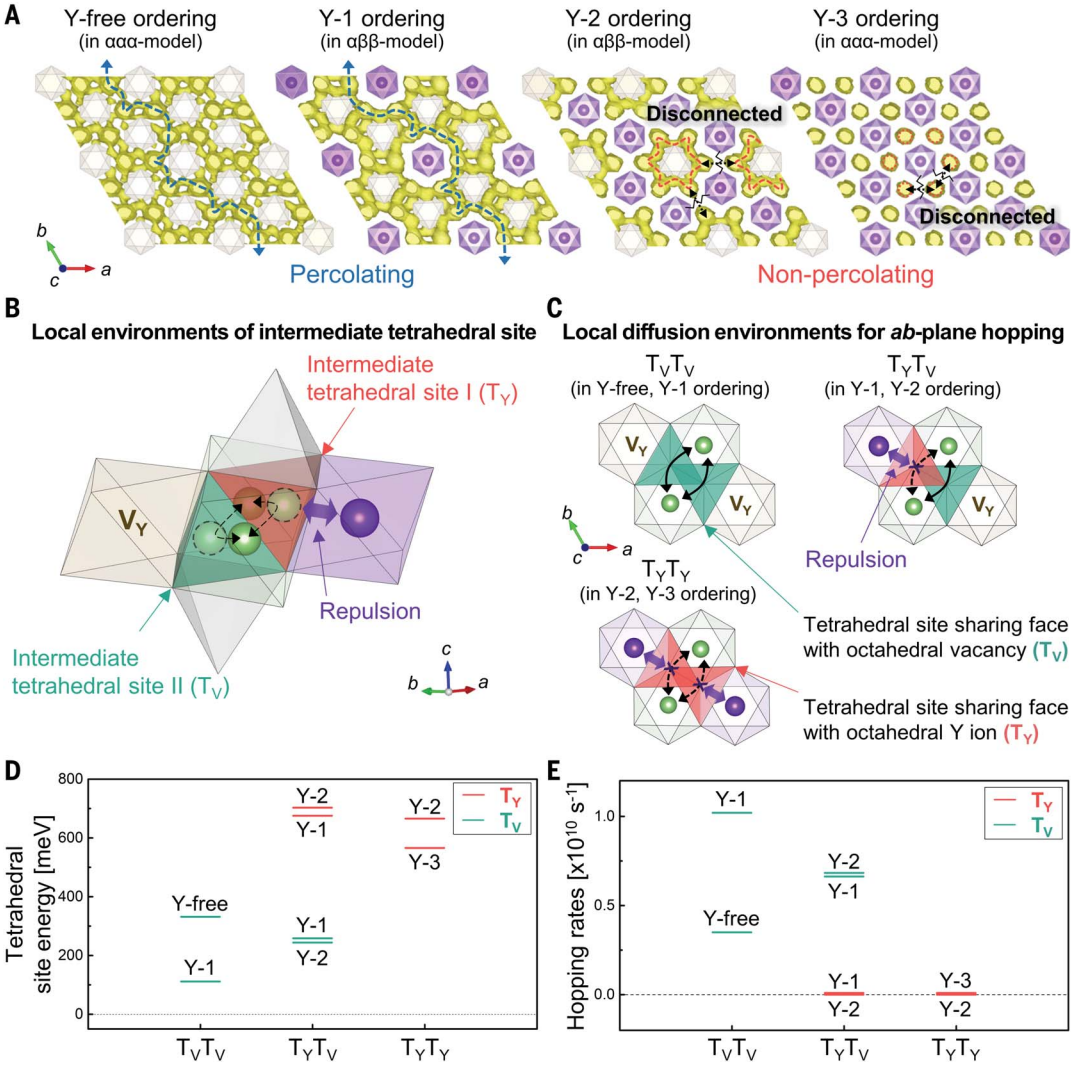


Fig. 2. Effect of cation arrangements on the percolating diffusion network. (**A**) Isosurface of lithium probability density $P = P_{max}/1024$ at 500 K in each *ab*-plane ordering (Y-free, Y-1, Y-2, and Y-3 orderings) to analyze the effects of the cation arrangements. Each *ab*-plane ordering can be divided into two groups, percolating (Y-free and Y-1) and nonpercolating (Y-2 and Y-3). A disconnected *ab*-plane lithium-diffusion pathway was observed in the specific diffusion pathway between Y ions marked by the black arrows. (**B**) Local environments of intermediate tetrahedral sites in the structure of Li_3YCl_6. There are two diffusion paths and their related intermediate sites (I and II). Each tetrahedral site shares faces with two Li or vacancy octahedrons and one Y or vacancy octahedron in the same *ab* plane but does not share any faces with any cation octahedrons in the layer above or below. Intermediate site I (red) shares a face with one Y octahedron, whereas intermediate site II (green) shares a face with a vacancy; intermediate sites I and II are denoted by T_Y and T_V, respectively. (**C**) Three local diffusion environments in the *ab* plane according to the types of the two intermediate tetrahedral sites. (**D**) Relative site energy of the lithium ion in the transition-state tetrahedral site. (**E**) Hopping rates for the three local diffusion environments. The green and red bars correspond to the calculation results of T_V and T_Y, respectively.

of six lithium ions, implying that sole diffusion in the *ab* plane cannot contribute to overall conductivity.

The in-plane percolation network loss was carefully examined by probing the energetics of lithium hopping in each Y ordering. In the in-plane diffusion, lithium ions migrate between octahedral sites via an intermediate tetrahedral site, analogous to the tetrahedral site hopping mechanism in well-known layered transition-metal oxides (*16*, *19*). Although the tetrahedral site in the LYC structure does not face-share with any cation in adjacent layers, it neighbors the Y octahedron in the same layer, electrostatically destabilizing the migrating lithium (*20*) (Fig. 2B). Consequently, the activation barrier in the hcp-stacked LYC structure can be approximated by the lithium state in the intermediate tetrahedral site, consistent with findings in O3-type layered lithium transition-metal oxides (*17*, *21*) (supplementary text, section 3). Depending on Y site occupancy, migrating lithium in the transition state may share a face with the Y ion in the site (T_Y) or with a vacant site (T_V). The Y-1 and Y-2 layers possess T_VT_V and T_YT_V paths for lithium migration, whereas the Y-free and Y-3 layers can only have T_VT_V and T_YT_Y paths, respectively (Fig. 2C).

Figure 2D plots lithium hopping transition-state energies for each path in the corresponding Y orderings. This figure demonstrates that low energy barriers are required for the paths passing by T_V (green bar), provided by the T_VT_V and T_YT_V environments in the Y-free, Y-1, and Y-2 layers. By contrast, paths involving T_Y (red bar) exhibit considerably higher energy barriers, indicating substantially subdued lithium mobility in T_YT_Y or T_YT_V environments. This finding suggests that the Y-3 layer, with only the T_YT_Y path, would not permit facile lithium conduction, consistent with Fig. 2A and fig. S4. To further quantify diffusion kinetics, lithium hopping

rates of each path were evaluated from AIMD results at 500 K using the analytical method developed by de Klerk *et al.* (*22*) (see methods). The results manifest that transition-state energies correlate with hopping rates, with more frequent hopping events for T_V sites in T_VT_V or T_YT_V environments (Fig. 2E). Hopping rates for T_VT_V (0.35×10^{10} and 1.02×10^{10} s^{-1}) and T_YT_V (0.67×10^{10} s^{-1}) paths are approximately 1000 times higher than those for T_YT_Y (1.5×10^{7} s^{-1}) paths (supplementary text, section 4).

Pillar effect of cations in the interlayer space

Intriguingly, the Y-1 layer exhibited a lithium hopping rate that was more than twice that of the Y-free layer for the same T_VT_V path. Additionally, hopping events in the Y-1 and Y-2 layers along the T_YT_V path are more frequent than in the Y-free layer through the T_V site. This indicates that lithium mobility tends to be higher in layers with some Y occupancy, which is counterintuitive because transition metals in the lithium layer (e.g., cation disorder) typically inhibit lithium diffusion by contracting the interlayer space in conventional layered lithium transition-metal oxides (*23*, *24*). To clarify this observation, we conducted a nudged elastic band calculation, comparing activation barriers for the T_VT_V path in Y-free and Y-1 layers (fig. S6), which revealed a 50-meV lower barrier in the Y-1 layer.

To investigate the origin of this unexpected phenomenon, we examined several potential factors that affect the stability of the transition-state lithium in the T_VT_V path (*23*, *25*), suspecting a change in interlayer distance as a determinant, as seen in our previous layered-oxide studies (*17*, *23*). In the layered framework, transition-state lithium stability primarily depends on interlayer space, with lithium-ion mobility generally increasing with interlayer distance. We discovered a similar relationship in the LYC halide, with the activation barrier for lithium hopping decreasing linearly with interlayer space, regardless of Y-ordering type (Fig. 3A). Furthermore, Fig. 3A reveals that the lower activation barrier in the Y-1 layer results from a larger equilibrium interlayer space (3.10 Å, green dotted line) in the Y-1 layer than in the Y-free layer (2.99 Å, red dotted line). This inverse correlation between interlayer distance and activation barrier is also valid for the T_YT_V path (fig. S7), signifying a similar lithium-diffusion mechanism to conventional layered lithium transition-metal oxides.

To account for the large interlayer space of the Y-1 layer, we systematically investigated the local environment of Y and calculated the shortest distance between Y ions to estimate the electrostatic repulsive effect. Figure 3B shows that the distance between neighboring Y ions becomes markedly shorter in the Y-1 layer (7.19 Å) than in the Y-free layer (8.83 Å), leading to a stronger *c*-axis repulsion component between Y ions in adjacent planes, which contributes to an increased interlayer space from 5.97 to 6.17 Å (*26*). The repulsion component along the *c* axis was further quantified by artificially altering Y occupancy in the interlayer and examining its correlation with interlayer distance (Fig. 3C) (supplementary text, section 5). This graph shows that as Y ions occupy the interlayer and increase electrostatic repulsion, the interlayer distance increases. The distance increases to 3.10 Å at a Y occupancy of 0.167 and then converges, indicating that Y ions serve as electrostatic pillars to maintain interlayer space (*27*). This highlights the importance of Y

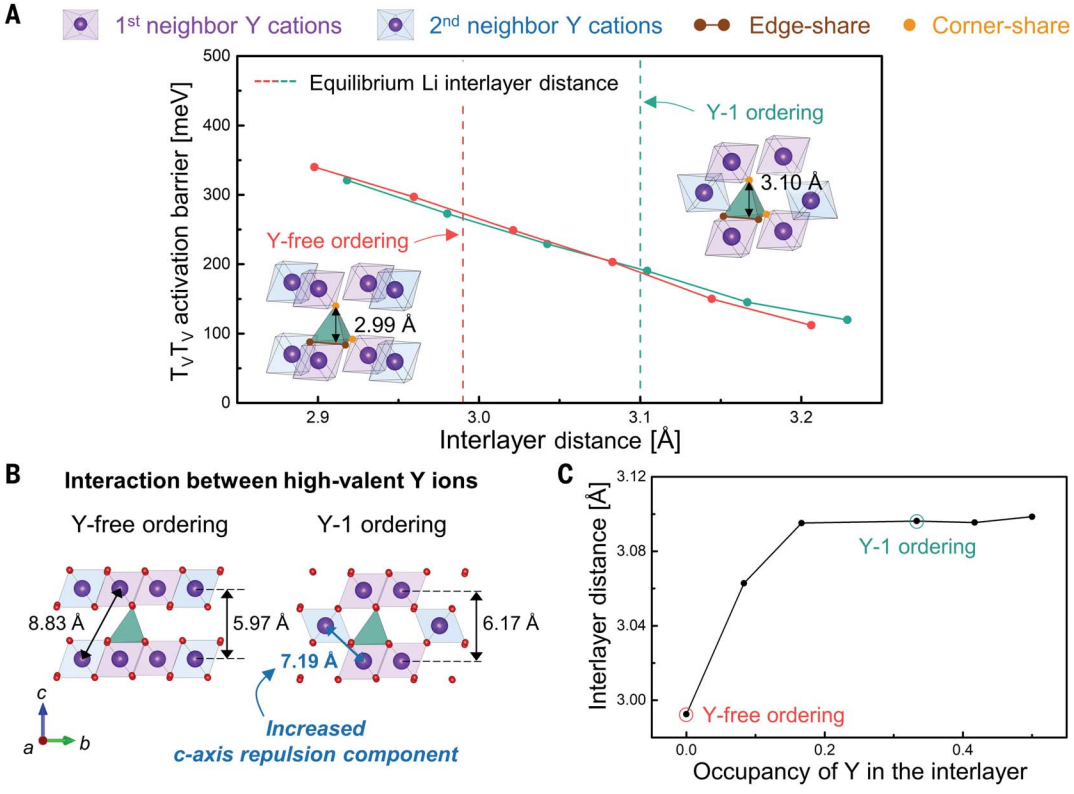


Fig. 3. Pillar effect of cations in the interlayer space. (**A**) T_VT_V activation barrier for *ab*-plane lithium diffusion in the Y-free (red) and Y-1 (green) orderings as a function of interlayer distance. The dashed lines indicate the calculated equilibrium Li interlayer distance of the Y-free (red) and Y-1 (green) orderings. The inset images show the local environment of the Y-free (left) and Y-1 (right) orderings. They have the same first-neighboring Y-ion local environment (purple octahedral site) but different second-neighboring Y-ion local environments (blue octahedral site). (**B**) Comparison of distances between Y ions from side views of Y-free (left) and Y-1 (right) orderings. (**C**) Interlayer distance as afunction of occupancy of Y $\left(\frac{\text{Number of } Y \text{ ions in the } ab \text{ plane}}{\text{Number of } Y \text{ ion sites in the } ab \text{ plane}}\right)$ in the interlayer. The Y-free and Y-1 orderings are denoted by red and green circles, respectively.

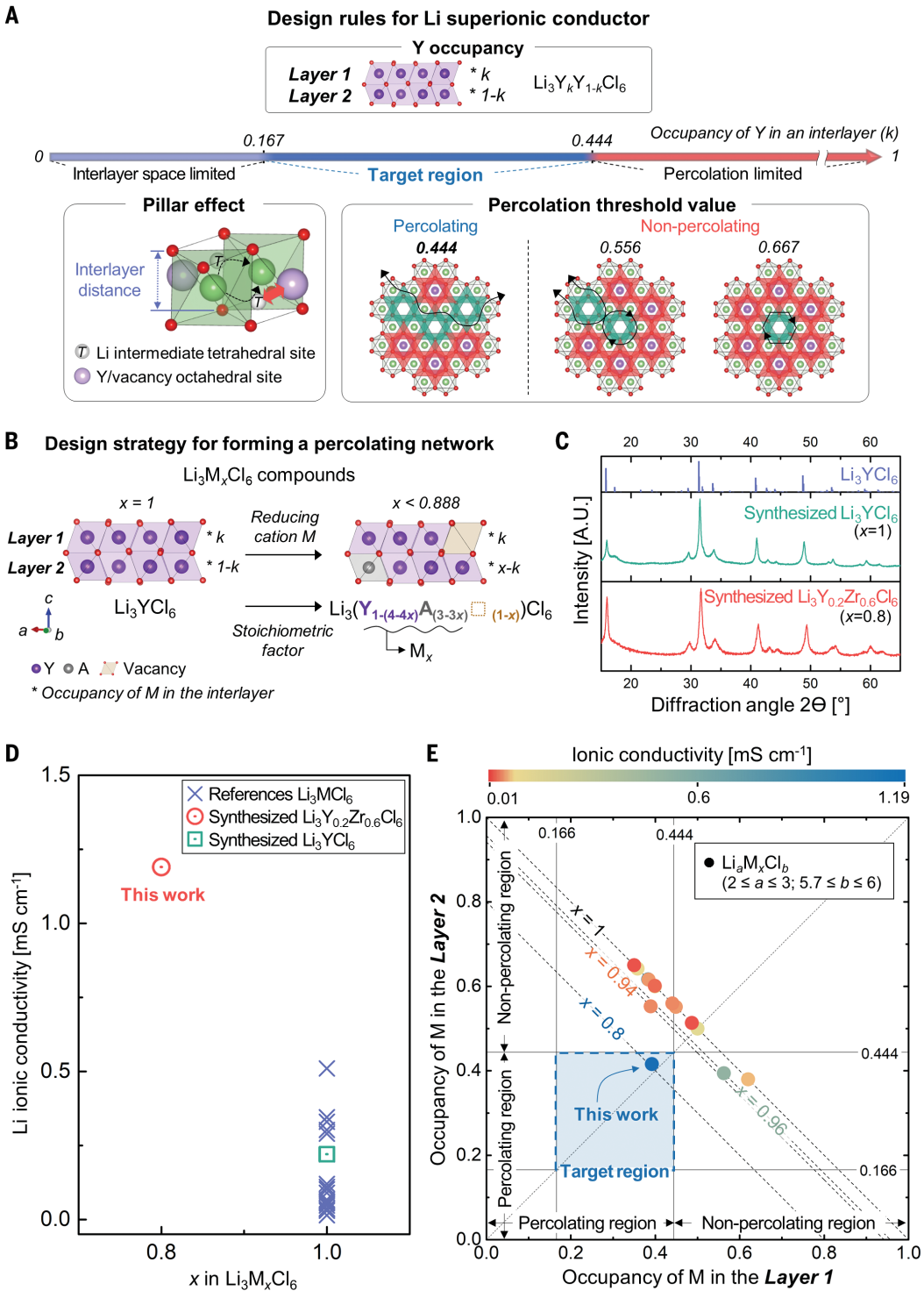


Fig. 4. Design of halide superionic conductor $Li_3Y_{0.2}Zr_{0.6}Cl_6$. (**A**) Proposed design rules of cation arrangements, which regulate two diffusion-determining factors in trigonal halide superionic conductors. (**B**) Design strategy that could construct percolating Li-diffusion pathways for all *ab* planes in trigonal halide superionic conductors. (**C**) Measured XRD patterns of mechanochemically synthesized Li_3YCl_6 (green) and $Li_3Y_{0.2}Zr_{0.6}Cl_6$ (red). The reference XRD pattern of Li_3YCl_6 (space group $P\bar{3}m1$) is shown in the top panel (*7*). A.U., arbitrary units. (**D**) Measured ionic conductivities of $Li_3Y_{0.2}Zr_{0.6}Cl_6$ and Li_3YCl_6 (empty circle and square, respectively) and of Li_3MCl_6 from previous works (*7*, *12*, *15*, *28*, *29*) (M indicates transition metal and is marked as X). (**E**) Ionic conductivities according to M cation occupancy of each *ab* plane, which are taken from the literature (*5*, *7*, *12*, *15*, *28*, *30*, *31*). Most of these are based on Li_3MCl_6 stoichiometry, except for that from (*5*), which is based on Li_2MCl_6 stoichiometry. Note that the refined compositions driven from the structure refinements from (*7*) are $Li_3Y_{0.957}Cl_{5.79}$ and $Li_{2.64}Y_{0.942}Cl_{5.82}$. The ionic conductivity value follows the color scale bar shown above the graph.

partial occupancy in expanding LYC halide interlayer space and facilitating *ab*-plane lithium diffusion, with the maximum interlayer space reached at a minimum partial occupancy of 0.167, which serves as a redesign criterion for LYC and is discussed later in the text (supplementary text, section 6).

Design of a halide superionic conductor

Our theoretical investigations revealed that lithium diffusion in trigonal halide conductors is governed by lithium percolating networks and interlayer space, which are both influenced by Y content and ordering. With a trade-off between the two factors regarding Y occupancy in the *ab* plane, establishing a suitable Y-ion arrangement is essential for enhancing ionic conductivity. To identify the ideal Y occupancy range, we first calculated the percolation threshold value of Y occupancy in the *ab* plane (Fig. 4A). In the trigonal symmetry of LYC, random Y distribution in 2d sites statistically results in a 0.444 Y-ion occupancy threshold, above which lithium path connectivity is locally confined, inhibiting lithium diffusion (supplementary text, section 7).

Although low Y occupancy benefits lithium percolation, reducing occupancy below a certain level leads to insufficient interlayer space, indicating that a minimum Y occupancy in the *ab* plane of nearly 0.167 is needed, which is shown in Fig. 3C. Moreover, a large interlayer space can be maintained when the Y occupancy reaches a 0.444 percolation threshold. Figure 4A shows the target Y occupancy range as between 0.167 and 0.444, which enables both lithium percolation networks and an enlarged interlayer space, resulting in fast lithium conduction (see supplementary text, section 8, for other symmetry of hcp-stacked Li_3MCl_6). However, the upper boundary (0.444) is unexpectedly low, considering that the total Y occupancy should be unity when adding partial Y occupancies (k) in two layers in stoichiometric $Li_3Y_kY_{1-k}Cl_6$ (top panel of Fig. 4A). Simply adding two layers of the ideal 0.444 occupancies cannot satisfy the target Li_3YCl_6 composition, resulting in $Li_3Y_{0.888}Cl_6$.

Instead, we propose an alternative strategy to lower M cation occupancy by partially substituting Y^{3+} ions with tetravalent ions in Li_3MCl_6 (Fig. 4B). Substituting trivalent ions with higher-valent ions (A^{4+}) would generate M cation vacancies without affecting the charge-carrying lithium-ion content (e.g., $Li_3Y_{4x-3}A_{3-3x}Cl_6$). When x is between 0.333 and 0.888, both layers 1 and 2 can theoretically fall within the region of 0.167 to 0.444. In this regard, we aimed for the maximum M occupancy of 0.888 and set the A^{4+} substitution level as $Li_3Y_{0.2}A_{0.6}Cl_6$ (or $Li_3M_{0.8}Cl_6$). Before further discussing this composition, we note that other approaches also can reduce M occupancy below unity, such as altering the lithium concentrations or using anion substitutions, which are expected to have similar effects. However, we take this partial substitution strategy as one of the accessible models of a proof-of-concept experiment. Accordingly, a trigonal halide compound, $Li_3Y_{0.2}A_{0.6}Cl_6$ (where A is Zr), was successfully synthesized by the mechanochemical route (see methods). The x-ray diffraction (XRD) patterns of synthesized $Li_3Y_{0.2}Zr_{0.6}Cl_6$ (Fig. 4C and fig. S13) resembled reported LYC (*7*), confirming the preservation of the trigonal LYC structure even with substantial Zr substitutions. More importantly, Rietveld refinement of the high-resolution powder diffraction (HRPD) analysis revealed that partial M occupancies in layers 1 and 2 are 0.392 and 0.416, respectively, which are both within the target range of 0.167 to 0.444 (fig. S14 and table S2). This validates that Zr substitution enables the generation of vacant M sites that are necessary for the lithium percolating network (supplementary text, section 9).

Subsequently, the ionic conductivity of $Li_3Y_{0.2}Zr_{0.6}Cl_6$ was measured and compared with that reported for LYCs (*7*, *12*, *15*, *28*, *29*) from various synthetic routes, including our own experiment (Fig. 4D and fig. S16). $Li_3Y_{0.2}Zr_{0.6}Cl_6$ demonstrates a conductivity of 1.19 mS cm^{-1}, which is almost six times higher than that of stoichiometric LYC (~0.2 mS cm^{-1}), despite sharing the same charge-carrier concentration and crystal structure. This value is the highest among trigonal halide–based SEs with $Li_3M_xCl_6$ stoichiometry (where M is a transition metal; $0 < x \leq 1$). It is also noteworthy that this simple design principle, based on target M partial occupancy, could be further validated with other trigonal halide electrolytes. In this regard, we revisited all the reported trigonal halide SEs, determining M cation occupancies in layers 1 and 2 from their structural data as well as the ionic conductivities, and plotted them in Fig. 4E (*5*, *7*, *12*, *15*, *28*, *30*, *31*). The results demonstrate that none of the materials fall in the target region (blue dashed square) except for $Li_3Y_{0.2}Zr_{0.6}Cl_6$, which implies that regulating the cation arrangements is crucial for improving the ionic conductivity (supplementary text, section 10).

Many works have reported that ball-milled LYCs exhibit higher ionic conductivities than annealed LYCs (*7*, *12*, *15*). This difference has been simply attributed to lattice expansion or structural distortion, which can open up lithium-diffusion paths after high-energy ball-milling. However, our revisit of these works and their structural data indicates that the difference stems from the interplay between altered Y occupancies that result from ball-milling synthesis and *ab*-plane lithium mobility. We found that annealed LYCs generally do not contain any *ab* planes with Y composition in the target region, whereas ball-milled LYCs mostly contain at least one *ab* plane with Y composition in the target region (table S1). This finding implies that sluggish in-plane lithium diffusion is expected in both *ab* planes (i.e., layers 1 and 2) of annealed LYCs. By contrast, achieving higher ionic conductivity is possible with ball-milled LYCs because of the presence of a plane exhibiting a percolating diffusion network with an enlarged interlayer space. These diffusion mechanisms align with findings from previous studies on conventional layered lithium transition-metal oxides and state-of-the-art disordered rocksalt cathodes, which both share structural similarities with trigonal halide structures, highlighting the importance of interlayer space (*17*, *23*) and percolation (*25*, *32*) for lithium diffusion. Therefore, our findings suggest that securing at least one *ab* plane in the target region is indispensable for achieving high ionic conductivity in the stoichiometric Li_3MCl_6 SEs, although this condition alone cannot fully satisfy our cation-arrangement criteria.

Conclusions

We elucidated the detailed superionic conduction mechanism of halide SEs and demonstrated an innovative strategy to enhance ionic conductivity by establishing the correlation between cation arrangements and lithium-diffusion kinetics. We found that the percolating diffusion network and interlayer distance are key factors that determine the lithium superionic conduction in trigonal hcp-stacked Li_3MCl_6 and are governed by the cation (M) arrangements. Although M serves as a pillar to enlarge the interlayer space, thereby promoting in-plane lithium diffusion, that is, the bottleneck in the overall diffusivity, excessive M content disrupts the lithium percolation network in the plane. This trade-off relationship suggests that boundary conditions for M occupancy in the *ab* plane exist for optimal lithium ionic conductivity, which serves as a design rule. As a proof-of-concept experiment, we introduced $Li_3Y_{0.2}Zr_{0.6}Cl_6$, a halide superionic conductor that has percolating diffusion networks with enlarged interlayer space and thus exhibits enhanced ionic conductivity.

REFERENCES AND NOTES

1. Y.-S. Hu, *Nat. Energy* **1**, 16042 (2016).
2. R. Chen, Q. Li, X. Yu, L. Chen, H. Li, *Chem. Rev.* **120**, 6820–6877 (2020).
3. X. Li *et al.*, *Energy Environ. Sci.* **13**, 1429–1461 (2020).
4. S. Wang *et al.*, *Angew. Chem. Int. Ed.* **58**, 8039–8043 (2019).
5. K. Wang *et al.*, *Nat. Commun.* **12**, 4410 (2021).
6. W. D. Richards, L. J. Miara, Y. Wang, J. C. Kim, G. Ceder, *Chem. Mater.* **28**, 266–273 (2016).
7. T. Asano *et al.*, *Adv. Mater.* **30**, e1803075 (2018).
8. K. Park *et al.*, *Chem. Mater.* **28**, 8051–8059 (2016).
9. J. C. Bachman *et al.*, *Chem. Rev.* **116**, 140–162 (2016).
10. H. Lutz, W. Schmidt, H. Haeuseler, *J. Phys. Chem. Solids* **42**, 287–289 (1981).
11. X. Li *et al.*, *Energy Environ. Sci.* **12**, 2665–2671 (2019).
12. R. Schlem *et al.*, *Adv. Energy Mater.* **10**, 1903719 (2020).
13. Y. Han *et al.*, *Adv. Energy Mater.* **11**, 2100126 (2021).
14. H. J. Steiner, H. Lutz, *Z. Anorg. Allg. Chem.* **613**, 26–30 (1992).
15. R. Schlem, A. Banik, S. Ohno, E. Suard, W. G. Zeier, *Chem. Mater.* **33**, 327–337 (2021).
16. D. Park *et al.*, *ACS Appl. Mater. Interfaces* **12**, 34806–34814 (2020).

17. K. Kang, Y. S. Meng, J. Bréger, C. P. Grey, G. Ceder, *Science* **311**, 977–980 (2006).
18. N. Phattharasupakun *et al*., *J. Electrochem. Soc.* **168**, 090535 (2021).
19. A. Van der Ven, G. Ceder, *J. Power Sources* **97-98**, 529–531 (2001).
20. J. Reed, G. Ceder, *Chem. Rev.* **104**, 4513–4533 (2004).
21. A. Van der Ven, J. Bhattacharya, A. A. Belak, *Acc. Chem. Res.* **46**, 1216–1225 (2013).
22. N. J. J. de Klerk, E. van der Maas, M. Wagemaker, *ACS Appl. Energy Mater.* **1**, 3230–3242 (2018).
23. K. Kang, G. Ceder, *Phys. Rev. B* **74**, 094105 (2006).
24. H. Yu *et al*., *Energy Environ. Sci.* **7**, 1068–1078 (2014).
25. J. Lee *et al*., *Science* **343**, 519–522 (2014).
26. D. Eum *et al*., *Nat. Mater.* **19**, 419–427 (2020).
27. R. A. House *et al*., *Nat. Energy* **6**, 781–789 (2021).
28. H. Ito *et al*., *Adv. Sci.* **8**, e2101413 (2021).
29. X. Li *et al*., *Nano Lett.* **20**, 4384–4392 (2020).
30. S. Y. Kim *et al*., *ACS Mater. Lett.* **3**, 930–938 (2021).
31. J. Park *et al*., *Chem. Eng. J.* **425**, 130630 (2021).
32. J. Huang *et al*., *Nat. Energy* **6**, 706–714 (2021).

ACKNOWLEDGMENTS

HRPD analysis was performed at the 9B beamline of the Pohang Light Source at the Pohang Accelerator Laboratory. **Funding:** This work was supported by a National Research Foundation of Korea grant funded by the Korean government (2021M3H4A1A04093050) (K.K.). This work was also supported by the research program of LOTTE Energy Materials (K.K.).
Author contributions: Conceptualization: S.Y.; Methodology: S.Y., J.N., B.Ki., J.-H.S., K.O., J.Y., S.L., S.-O.P., W.K., B.Ka., D.K.; Investigation: S.Y., J.N., B.Ki., J.-H.S., S.L.; Visualization: S.Y., J.N., B.Ki., J.-H.S.; Funding acquisition: J.-H.S., K.K.; Project administration: K.K.; Supervision: K.K.; Writing – original draft: S.Y., J.N., B.Ki., J.-H.S., K.K.; Writing – review and editing: S.Y., J.N., K.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg6591
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 to S3
References (*33–50*)

Submitted 12 January 2023; accepted 4 September 2023
10.1126/science.adg6591

WATER AVAILABILITY

Southern Hemisphere dominates recent decline in global water availability

Yongqiang Zhang[1]*†, Congcong Li[2,3]†, Francis H. S. Chiew[3], David A. Post[3], Xuanze Zhang[1], Ning Ma[1], Jing Tian[1], Dongdong Kong[4], L. Ruby Leung[5], Qiang Yu[2], Jiancheng Shi[6], Changming Liu[1]

Global land water underpins livelihoods, socioeconomic development, and ecosystems. It remains unclear how water availability has changed in recent decades. Using an ensemble of observations, we quantified global land water availability over the past two decades. We show that the Southern Hemisphere has dominated the declining trend in global water availability from 2001 to 2020. The significant decrease occurs mainly in South America, southwestern Africa, and northwestern Australia. In the Northern Hemisphere, the complex regional increasing and decreasing trends cancel each other, resulting in a negligible hemispheric trend. The variability and trend in water availability in the Southern Hemisphere are largely driven by precipitation associated with climate modes, particularly the El Niño-Southern Oscillation. This study highlights their dominant role in controlling global water availability.

Water availability over the global land is the net difference between water supply from precipitation and water demand from evapotranspiration (ET), thus representing streamflow and water storage change. Global land water availability influences livelihoods, socioeconomic development, and ecosystems (*1*–*3*) and could dramatically shift with climate change and socioeconomic growth as projected over the next several decades (*4*, *5*). In the past several decades, water availability has varied because of climate change that has intensified the hydrological cycle, vegetation greening that has increased terrestrial ET, and human water use that has altered streamflow (*6*). However, global land surface drought has changed little in the same period (*7*, *8*) because of complementary trends in precipitation and ET, which are controlled by multiple climate factors, such as available energy, humidity, and wind speed (*9*, *10*). Human activities have played a strong role in altering water availability by decreasing groundwater and streamflow, with potentially devastating effects on aquatic ecosystems and food production (*11*–*13*). Human-managed reservoirs have contributed 57% of the change in Earth's seasonal surface water storage variability. Human activities now drive 67% of the variability in surface water storage in the northern midlatitudes and nearly 100% in arid and semiarid regions (*5*). The observed trend in river flows can be explained only if anthropogenic climate change impact is included. Humans also influence the climate and affect the magnitude of river flows at regional and global scales (*14*). More than 70% of global net permanent water loss occurs in the Middle East and Central Asia, which is linked to drought and human activities, including damming, river diversion, and unregulated withdrawal (*15*). Global land surface could become drier because of the increasing trend in ET (*16*), particularly in the Northern Hemisphere, where vegetation greening (*17*) has controlled the decadal changes in ET (*18*, *19*). However, a recent study showed that vegetation greening could also increase water availability for approximately 45% of the global land surface (*20*) because the resulting moisture is recycled from vegetation to precipitation.

The above discussion highlights the conflicting conclusions that different studies have drawn about changes in water availability; there are several reasons for this. First, different periods are used in the different studies, varying from one to six decades (*1*, *7*, *21*, *22*). Second, different methods are used to estimate water availability changes; for example, some studies use water storage changes from Gravity Recovery and Climate Experiment (GRACE) satellite observations to reflect changes in water availability (*22*), whereas other studies use the climate drought severity index, potential ET (*7*, *23*, *24*), or river flow (*14*, *25*, *26*) to indicate water availability changes. Without large-scale validations against observations across the global land surface, water availability changes estimated using different approaches are questionable.

In this study, we analyzed the past two decades to elucidate changes in water availability across the globe and their causality. We validated the water availability trend using a consistent method against GRACE satellite and streamflow observations. We used precipitation minus ET (P – ET) to estimate water availability on the land surface and assumed that the trend in P – ET in a river basin is identical to the trend of water storage change plus streamflow at monthly to decadal scales. We combined multiple remote sensing and ground-based precipitation and ET datasets using simple model output averaging (SMA), Bayesian model averaging (BMA), and machine learning (ML) and developed an ensemble estimate for this study (materials and methods). The ensemble mean annual P – ET trend compares well [coefficient of determination (R^2) = 0.81] with the trend observed (water storage change plus streamflow) from 2003 to 2016 in 134 river basins that cover 36% of the global land surface (figs. S1 to S6 and table S1).

Interannual variability and trend

The annual anomaly of water availability across global land surface from 2001 to 2020 varies from −30 mm year^{-1} to +22 mm year^{-1}, with an annual variance of 146 mm^2 year^{-2} (Fig. 1A). The interannual variability in global water availability is largely contributed by the Southern Hemisphere. Despite accounting for only 26% of the global land area (not including Antarctica), the Southern Hemisphere contributes 43 ± 2.6% of the total global water availability interannual variability with its variance and 19 ± 3.9% with the covariance between the two hemispheres removed. The Northern Hemisphere, accounting for 74% of the global land area, contributes 38 ± 2.4% of the global water availability variance.

The global land water availability declined from 2001 to 2020, with a trend of −0.96 ± 0.27 mm year^{-2} ($P < 0.05$, table S2). The water availability across the Southern Hemisphere decreased significantly (−3.55 ± 0.71 mm year^{-2}, $P < 0.05$), but there was little trend in water availability across the Northern Hemisphere (−0.05 ± 0.12 mm year^{-2}). The percentage contributions from the Southern Hemisphere and the Northern Hemisphere are 95.2 ± 5.6% and 4.8 ± 5.6%, respectively. The Southern Hemisphere therefore dominated the global declining water availability trend from 2001 to 2020, despite its much smaller land surface area (Fig. 1, B and C). Over the Northern Hemisphere, water availability increased in southern China, eastern North America, and northeastern Asia and decreased in mainland southeast Asia, eastern Europe, and large parts of Siberia (Fig. 1B and fig. S7). The increasing and decreasing water availability trends in the different regions cancel each other, resulting in little trend when aggregated over the Northern Hemisphere. In the Southern Hemisphere, a strong decreasing trend is observed in South America ($P < 0.01$), the majority of Africa, and central and northwestern Australia ($P < 0.05$) (Fig. 1B). Nevertheless, there are some regions, such as the southern part of South America, that show a significant increasing trend ($P < 0.05$).

[1]Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China. [2]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling 712100, China. [3]CSIRO Environment, Black Mountain, Canberra, ACT 2601, Australia. [4]Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan, China. [5]Pacific Northwest National Laboratory, Richland, WA 99352, USA. [6]National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China.
*Corresponding author. Email: zhangyq@igsnrr.ac.cn
†These authors contributed equally to this work.

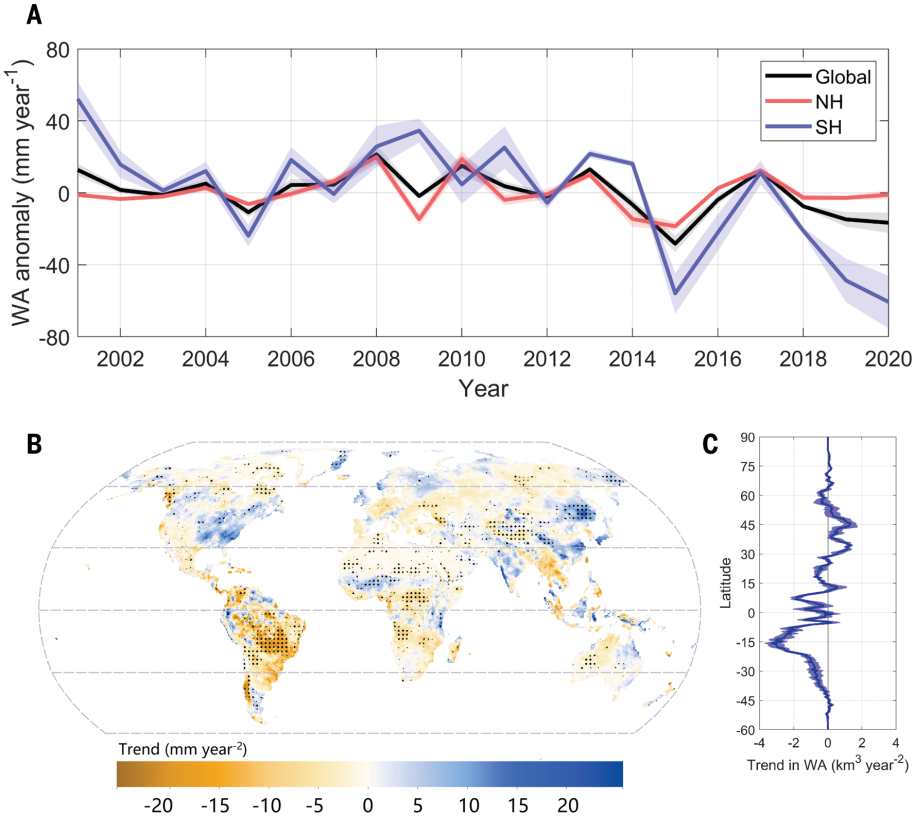


Fig. 1. Annual time series, variability, and trend of global land water availability (WA). (**A**) Annual time series of land WA across the globe (black), Northern Hemisphere (NH, red), and Southern Hemisphere (SH, blue). (**B**) Trend in WA at each 0.5° by 0.5° grid cell across the globe from 2001 to 2020. Stippling indicates that the trend is statistically significant at $P < 0.05$. (**C**) Aggregated trend in WA along each 0.5° latitude band (km^3 year^{-2}). The solid line and shading denote the mean and 1 SD of three estimates [simple model output (SMA) averaging, Bayesian model averaging, and machine learning], respectively.

Attribution and causality

The interannual variability of water availability in about 50% of the global land surface is dominated by precipitation (with a relative contribution ratio of variance in water availability > 0.6) (fig. S8). This is also the case for the trend in water availability in wet regions such as eastern China, northeastern Asia, Amazonia, eastern North America, and Europe (Fig. 2A and fig. S9). Conversely, in very dry regions such as the Sahara, central Asia, central Australia, and western United States, the trend in water availability is driven by the trend in ET or in both ET and precipitation (Fig. 2A and fig. S9). The spatial pattern of Fig. 2A is quantitively analyzed by using the correlation between the precipitation contribution (or ET contribution) and aridity index (Fig. 2, B and C), with correlation coefficient (r) = 0.41 ($P < 0.001$) and r = 0.33 ($P < 0.001$) across the land surface of the Northern Hemisphere and Southern Hemisphere, respectively. This analysis demonstrates the tendency for a higher precipitation contribution in wetter grid cells with a higher aridity index (Fig. 2, B and C).

The relationship between the precipitation contribution and the aridity index is further supported by the water availability trends in 40 large river basins (Fig. 2, D and E). Precipitation dominates the water availability trend in wetter basins such as the Saint Lawrence, Amazon, Mississippi, and Yangtze, whereas ET dominates the water availability trend in drier basins such as the Ob, Colorado, Okavango, and Tarim. Both precipitation and ET contribute to the water availability trend in some basins, such as the Yellow, Indus, Ganges, and Mekong (Fig. 2D). Strong trends in water availability are generally observed in basins where either precipitation or ET dominates the trend, and little trend is found in basins where precipitation and ET contribute more evenly. Both positive and negative water availability trends are observed where precipitation dominates, but the water availability trend is generally negative where ET dominates (Fig. 2, D and E).

There is a strong positive correlation between the annual water availability and the Southern Oscillation Index (SOI) in southeast Asia, the equatorial Pacific, Australia, central America, and northern South America (Fig. 3A). The SOI is the difference between surface air pressure in Tahiti and Darwin. Positive SOI values indicate La Niña conditions, and hence, a positive correlation indicates wetter conditions in the aforementioned regions during La Niña and drier conditions during El Niño. The water availability-El Niño-Southern Oscillation (ENSO) correlation here has also been observed for individual water cycle components, such as precipitation (*27*, *28*) and river flow (*29*).

In the Southern Hemisphere, regions with positive correlations between water availability and SOI largely show decreasing or insignificant water availability trends from 2001 to 2020 (Fig. 3B). By contrast, areas with negative water availability-SOI correlations display a tendency for increasing water availability trends in both the Northern and Southern Hemispheres (Fig. 3B).

There is no clear relationship between the water availability trend and the aridity index in the Northern Hemisphere (Fig. 3C). In the Southern Hemisphere, there is a relatively weak but statistically significant correlation between the water availability trend and the aridity index, with larger decreasing trends in wetter regions (r = −0.19, $P < 0.001$) (Fig. 3C). This observed "wet gets drier" water availability change from 2001 to 2020 is opposite to the "dry gets drier, wet gets wetter" hydrological change postulated under global warming.

There were significant increases ($P < 0.05$) in the leaf area index (LAI) or greening over large areas of the Northern Hemisphere from 2001 to 2020, especially in eastern China, India, and Europe (figs. S10 and S11). To investigate whether greening plays a role in the water availability and aridity index relationship, we separated the regions with increasing or decreasing LAI trends (Fig. 3C). In the Northern Hemisphere, there is no relationship between the water availability trend and the aridity index, and greening has little or no influence on the relationship. In the Southern Hemisphere, the statistically significant relationship between the water availability trend and the aridity index is stronger in grid cells that show an increasing LAI trend (r = −0.27, $P < 0.001$) (Fig. 3C). These results suggest that the large-scale greening in some parts of the Northern Hemisphere was accompanied by both increases and decreases in water availability, depending on the region. In the Southern Hemisphere, greening has accentuated the decline in water availability driven by the decline in precipitation (Fig. 2A), likely because greening further reduces water availability by increasing ET.

Discussion and implication

This study demonstrates that the interannual variability and trends in global water availability from 2001 to 2020 are largely dominated

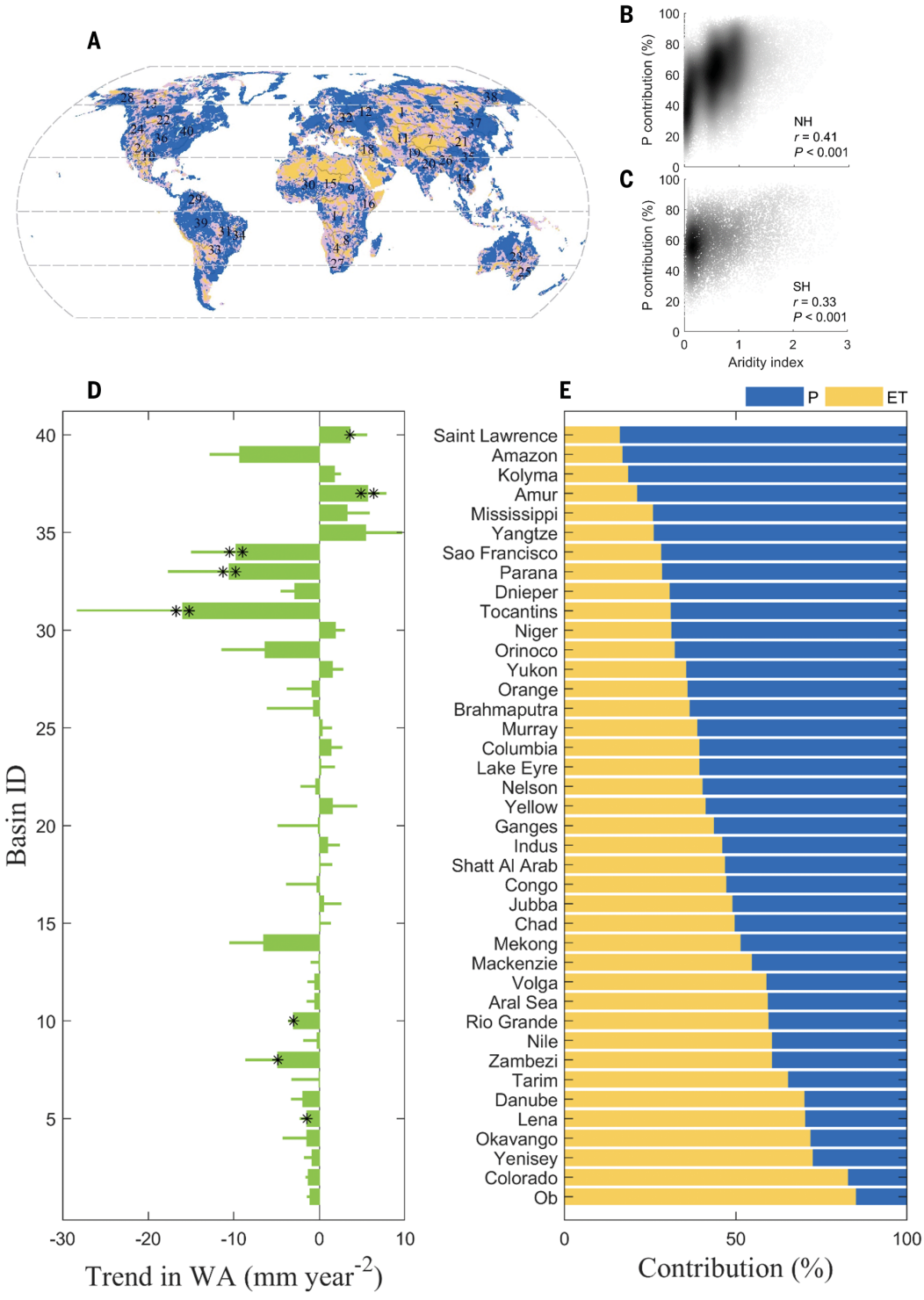


Fig. 2. Relative contribution of precipitation (P) and actual evapotranspiration (ET) to the trends in WA from 2001 to 2020. (**A**) Contribution across each grid cell. Blue indicates that P dominates (with a contribution larger than 60%), orange indicates that actual ET dominates (with a contribution larger than 60%), and purple means that both P and ET play an important role (both with a contribution of 40 to 60%). (**B**) Scatterplots of the P contribution to the water availability (WA) trend in each grid cell versus the aridity index (mean annual P divided by mean annual potential ET) for grid cells in the NH. A higher aridity index indicates a wetter climate. Darker regions have a higher density of points and cells. (**C**) Same as (B) except that the grid cells indicate the SH. (**D**) Aggregated trend for 40 large river basins, with each bar representing the mean (±1 SD) from nine estimates. *P < 0.1; **P < 0.05. (**E**) P and ET contributions to the WA trend in the 40 river basins. The results presented from (A) to (E) are for the SMA WA estimate (materials and methods).

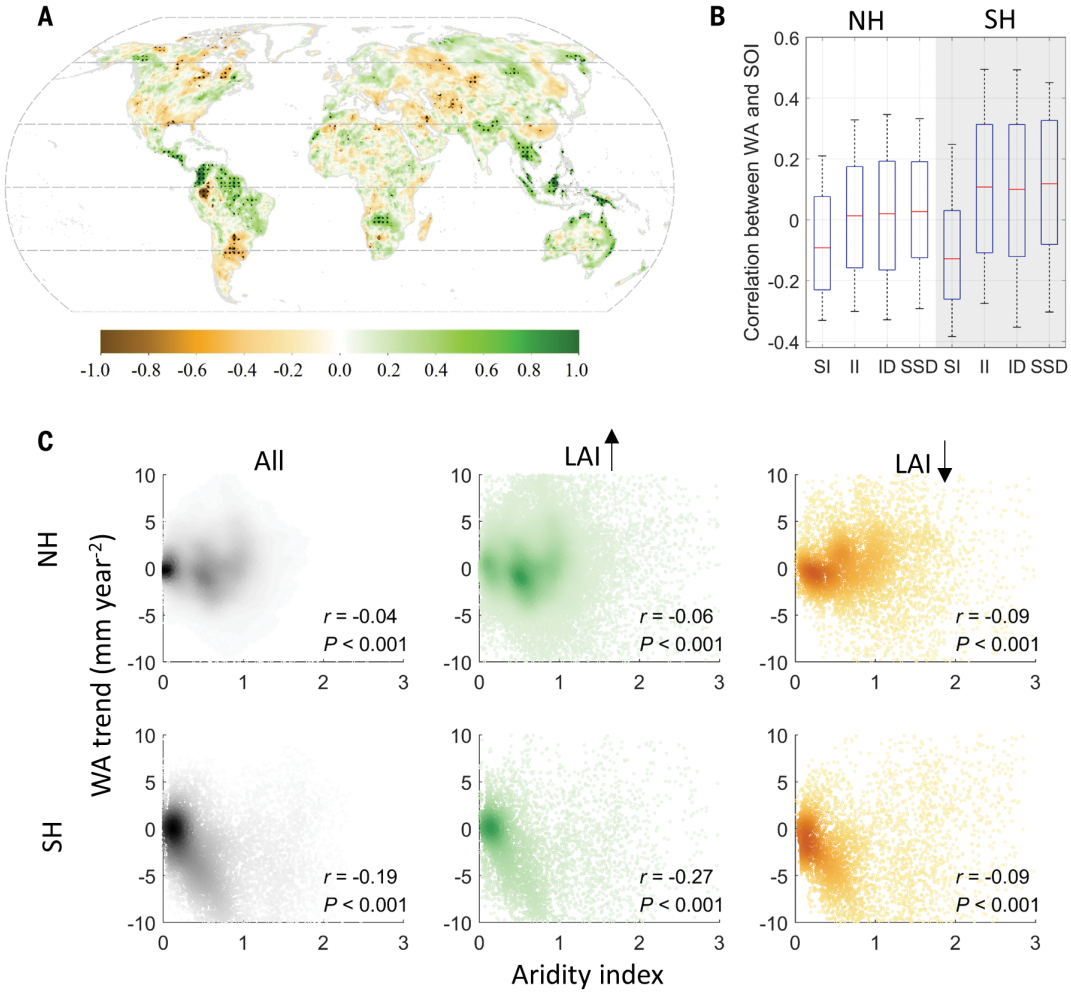


Fig. 3. Causality for interannual variability and trends in WA. (**A**) Correlation between annual WA and annual SOI from 2001 to 2020 at each 0.5° by 0.5° grid cell. Stippling indicates that the correlation is statistically significant ($P < 0.05$). (**B**) Correlation between WA and SOI grouped by different WA trend categories. For each boxplot, the bottom, middle, and top of the box are the 25th, 50th, and 75th percentiles, respectively, and the bottom and top whiskers show the 10th and 90th percentiles, respectively. SI, II, ID, and SSD indicate statistically significant increases in water availability ($P \leq 0.05$), nonstatistically significant increases in WA, nonstatistically significant decreases in WA, and statistically significant decreases in WA ($P \leq 0.05$), respectively, for the NH (white area) and the SH (gray area). (**C**) (Left) Scatterplots of trend versus aridity index for all grid cells, (middle) grid cells with increasing LAI, and (right) grid cells with decreasing LAI for the NH (top panels) and SH (bottom panels).

by the variability and trend in the Southern Hemisphere. The latter are largely dependent on the hydroclimate region and are linked to regional climate, natural variability of precipitation, and large-scale circulation drivers of changes in precipitation and ET, such as those related to the climate modes and vegetation greening (Fig. 4). The decline in water availability over large parts of the Southern Hemisphere in the relatively short 20-year period that we analyzed could be related to complex changes in climate modes and/or their relationships with precipitation, such as increasing ENSO-rainfall variability in a warming Earth (*27*, *30*, *31*) and changes in large-scale meridional circulations, such as the expanding Hadley cell (*32*, *33*). Among the various climate modes, ENSO is particularly important because it induces interannual variability in precipitation and water availability in large parts of the Southern Hemisphere and equatorial Pacific, with high water availability during La Niña and low water availability during El Niño. The ENSO roles are further corroborated by analyzing the correlations between water availability and the Niño 3.4 Index—another commonly used index (fig. S12)—and over a longer period from 1983 to 2020, confirming that the precipitation versus ENSO correlation in the Southern Hemisphere was stronger than that in the Northern Hemisphere (fig. S13). In addition, the expansion of the Hadley cell could be an important factor contributing to the drying in the Southern Hemisphere because this expansion pushes the winter storm tracks further south, away from the land masses and into the Southern Ocean (*33*, *34*). Notably, the Coupled Model Intercomparison Project Phase 5 (CMIP5) and CMIP6 models project a poleward shift of the Hadley cell edge that is about two to three times as large in the Southern Hemisphere as it is in the Northern Hemisphere (*35*).

The complex trends in water availability in the Northern Hemisphere can be partly related to direct and indirect human activities (Fig. 4). Irrigation can alter water availability regionally, but its role is secondary compared with natural climate variability and changes in precipitation and ET (*36*). For example, in the northern midlatitudes (16° to 22° and 35° to 42°), irrigation and greening may have contributed to the decreasing trend in water availability. Both LAI and water availability have largely increased, suggesting that vegetation greening may be

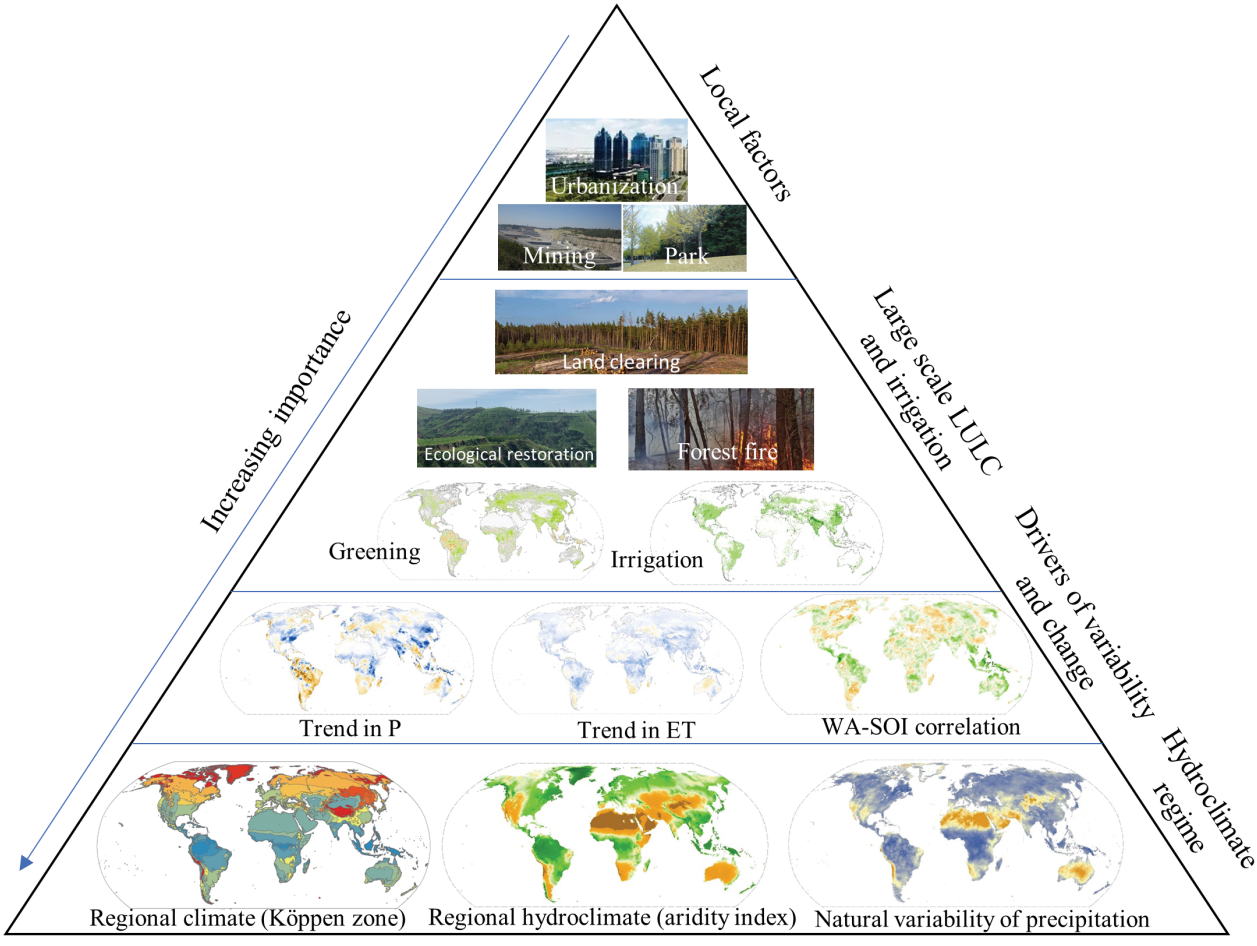


Fig. 4. Conceptualized understanding of major factors controlling variability and trends in WA. Natural variability of precipitation in the lower-right panel is expressed as the coefficient of variation of precipitation, the ratio of the SD of annual precipitation to the annual mean. LULC, land use and land cover.

positively coupled with regional water availability (*20*) (fig. S11). Therefore, greening could be a water availability driver (negative effect) and/or a response variable of water availability (positive effect), particularly in the northern midlatitudes, where both greening and water availability are strongly influenced by anthropogenic activities. Therefore, vegetation changes and large-scale irrigation can play a secondary role in controlling regional water availability (Fig. 4).

Our study has broad implications. First, it is crucial for water-resource managers to understand the dominant factors controlling interannual variability and trends in water availability. Precipitation has dominated the variability and trends in water availability, especially in the Southern Hemisphere, such as in South America and Australia, where river flow has been projected to continue to decline under climate change (*37*). Vegetation changes and irrigation, which are correlated in the irrigated regions, can play secondary roles at small to regional scales. Second, in terms of ecohydrology, vegetation greening could reduce local water availability, but our findings indicate that its large-scale impact is relatively weak. Vegetation greening can arise not only from an increase in precipitation but also from direct anthropogenic interventions, such as ecological restoration to mitigate land degradation and irrigation to enhance food production (*17*). Moreover, it fosters moisture recycling by increasing precipitation (*38*), indicating a coevolution of vegetation greening and water availability in the Northern Hemisphere. However, it is difficult to disentangle the interactions between vegetation and water availability despite recent efforts in modeling (*20*, *39*, *40*). Third, in many parts of the world, especially in wet regions such as the Amazon, ET plays a relatively minor role in determining water availability trends. Fourth, water availability provides a fundamental water source for carbon and nutrient cycles in ecosystems across the global land surface (*41*). Therefore, our results provide some indication for trends in vegetation growth (*42*), carbon, and other nutrients, including regional declining trends in the carbon cycle in Amazonia (*43*). Last, our results and conclusions pertain to the past two decades and thus can at least provide a constraint for global models driven by the observed sea surface temperature variability.

It remains challenging to determine and interpret water availability trends and their drivers over a longer period before 2000 because large basin-scale validation data are not available and because robust and consistent remotely sensed ET estimates are available only for the recent decades (*44*). Fusion of observations and global model simulations, including multimodel ensembles and large-ensemble simulations, can potentially reduce uncertainty in estimating water availability over a longer period.

REFERENCES AND NOTES

1. P. C. D. Milly, K. A. Dunne, A. V. Vecchia, *Nature* **438**, 347–350 (2005).
2. T. Oki, S. Kanae, *Science* **313**, 1068–1072 (2006).
3. M. Jung *et al.*, *Nature* **541**, 516–520 (2017).
4. Y. Pokhrel *et al.*, *Nat. Clim. Chang.* **11**, 226–233 (2021).
5. S. W. Cooley, J. C. Ryan, L. C. Smith, *Nature* **591**, 78–81 (2021).
6. G. Villarini, C. Wasko, *Nat. Clim. Chang.* **11**, 725–726 (2021).
7. J. Sheffield, E. F. Wood, M. L. Roderick, *Nature* **491**, 435–438 (2012).
8. P. Greve *et al.*, *Nat. Geosci.* **7**, 716–721 (2014).
9. X. Lian *et al.*, *Nat. Rev. Earth Environ.* **2**, 232–250 (2021).

10. T. R. McVicar *et al*., *J. Hydrol.* **416–417**, 182–205 (2012).
11. M. O. Cuthbert *et al*., *Nature* **572**, 230–234 (2019).
12. C. Dalin, Y. Wada, T. Kastner, M. J. Puma, *Nature* **543**, 700–704 (2017).
13. I. E. M. de Graaf, T. Gleeson, L. P. H. (Rens) van Beek, E. H. Sutanudjaja, M. F. P. Bierkens, *Nature* **574**, 90–94 (2019).
14. L. Gudmundsson *et al*., *Science* **371**, 1159–1162 (2021).
15. J. F. Pekel, A. Cottam, N. Gorelick, A. S. Belward, *Nature* **540**, 418–422 (2016).
16. Y. Zhang *et al*., *Sci. Rep.* **6**, 19124 (2016).
17. C. Chen *et al*., *Nat. Sustain.* **2**, 122–129 (2019).
18. Y. Yang *et al*., *Nat. Rev. Earth Environ.* **4**, 626–641 (2023).
19. C. Li, Y. Zhang, Y. Shen, Q. Yu, *Sci. Bull.* **65**, 1859–1861 (2020).
20. J. Cui *et al*., *Nat. Geosci.* **15**, 982–988 (2022).
21. M. Jung *et al*., *Nature* **467**, 951–954 (2010).
22. M. Rodell *et al*., *Nature* **557**, 651–659 (2018).
23. P. C. D. Milly, K. A. Dunne, *Nat. Clim. Chang.* **6**, 946–949 (2016).
24. A. Dai, *Nat. Clim. Chang.* **3**, 52–58 (2013).
25. Y. Yang *et al*., *Water Resour. Res.* **54**, 4700–4713 (2018).
26. J. S. Mankin, R. Seager, J. E. Smerdon, B. I. Cook, A. P. Williams, *Nat. Geosci.* **12**, 983–988 (2019).
27. A. Timmermann *et al*., *Nature* **559**, 535–545 (2018).
28. S. Power, F. Delage, C. Chung, G. Kociuba, K. Keay, *Nature* **502**, 541–545 (2013).
29. R. Emerton *et al*., *Nat. Commun.* **8**, 14796 (2017).
30. K.-S. Yun *et al*., *Commun. Earth Environ.* **2**, 43 (2021).
31. P. Huang, S.-P. Xie, *Nat. Geosci.* **8**, 922–926 (2015).
32. V. B. P. Chagas, P. L. B. Chaffe, G. Blöschl, *Nat. Commun.* **13**, 5136 (2022).
33. D. J. Seidel, Q. Fu, W. J. Randel, T. J. Reichler, *Nat. Geosci.* **1**, 21–24 (2008).
34. S.-Y. Kim *et al*., *Sci. Adv.* **9**, eadg1801 (2023).
35. K. M. Grise, S. M. Davis, *Atmos. Chem. Phys.* **20**, 5249–5268 (2020).
36. Z. Huang, X. Yuan, X. Liu, *J. Hydrol.* **601**, 126648 (2021).
37. Y. Zhang *et al*., *Nat. Water* **1**, 261–271 (2023).
38. R. J. van der Ent, H. H. G. Savenije, B. Schaefli, S. C. Steele-Dunne, *Water Resour. Res.* **46**, W09525 (2010).
39. S. Zhou *et al*., *Nat. Clim. Chang.* **11**, 38–44 (2021).
40. G. Miguez-Macho, Y. Fan, *Nature* **598**, 624–628 (2021).
41. A. Ahlström *et al*., *Science* **348**, 895–899 (2015).
42. W. Li *et al*., *Nat. Commun.* **13**, 3959 (2022).
43. L. V. Gatti *et al*., *Nature* **595**, 388–393 (2021).
44. Y. Zhang *et al*., *Remote Sens. Environ.* **222**, 165–182 (2019).

ACKNOWLEDGMENTS

Y.Q.Z. acknowledges support from the Chinese Academy of Sciences. C.C.L. thanks the Chinese Scholarship Council for financial support. F.H.S.C., D.A.P., and C.C.L. acknowledge support from the Australian Commonwealth Scientific and Industrial Research Organisation. The streamflow data used for the study were obtained from the Global Runoff Data Centre (GRDC), the US Geological Survey (USGS) National Water Information System and Geospatial Attributes of Gages for Evaluating Streamflow (GAGES)-II database, the Australian Bureau of Meteorology, and the *Chinese River Sediment Bulletin*. **Funding:** The study was supported by the National Key R&D Program of China (grant no. 2022YFC3002804), the National Natural Science Foundation of China (grant nos. 41971032 and 42330506), and the Second Tibetan Plateau Scientific Expedition and Research Program (2019QZKK020807 and 2019QZKK0206). **Author contributions:** Y.Q.Z. designed the study and wrote the first draft. C.C.L. collated most of the datasets and conducted the data analysis and model simulations. Y.Q.Z. and C.C.L. prepared the figures. F.H.S.C. and D.A.P. provided critical insights into the data analysis. N.M. contributed to the collation of the streamflow dataset. D.D.K. collated and processed the leaf area index dataset. All authors contributed to discussion, text revisions, and interpretations of the results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data and code are stored in the Figshare repository at https://doi.org/10.6084/m9.figshare.23949225.
License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh0716
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 and S2
References (*45*–*65*)

Submitted 8 February 2023; accepted 11 September 2023
10.1126/science.adh0716

THERMAL TRANSPORT

Electrically gated molecular thermal switch

Man Li[1]†, Huan Wu[1]†, Erin M. Avery[2,3], Zihao Qin[1], Dominic P. Goronzy[2,3], Huu Duy Nguyen[1], Tianhan Liu[2], Paul S. Weiss[2,3,4,5], Yongjie Hu[1,3,5]*

Controlling heat flow is a key challenge for applications ranging from thermal management in electronics to energy systems, industrial processing, and thermal therapy. However, progress has generally been limited by slow response times and low tunability in thermal conductance. In this work, we demonstrate an electronically gated solid-state thermal switch using self-assembled molecular junctions to achieve excellent performance at room temperature. In this three-terminal device, heat flow is continuously and reversibly modulated by an electric field through carefully controlled chemical bonding and charge distributions within the molecular interface. The devices have ultrahigh switching speeds above 1 megahertz, have on/off ratios in thermal conductance greater than 1300%, and can be switched more than 1 million times. We anticipate that these advances will generate opportunities in molecular engineering for thermal management systems and thermal circuit design.

Mastering electricity has driven the current age of modern information technology. For electronic devices, as early as the 1920s, scientists understood how to make a two-terminal device by contacting a sharp metal tip and a semiconductor crystal. These point-contact diodes have a large contrast in resistance depending on the flow direction of current and were used to rectify signals and make simple AM radio receivers. However, it took many years before the three-terminal solid-state device—the transistor—was invented at Bell Laboratories in 1947 (*1*). In the concept of electronics, the conductance can be continuously tuned by applying voltage through a third terminal. This breakthrough was recognized by the Nobel Prize in Physics in 1956 (*2*) and became a cornerstone of modern electronics, greatly influencing everyday life ever since.

For thermal devices, much work has been devoted to the control of heat flow for widespread applications ranging from thermal management in electronics to heating and cooling facilities in households and industry, thermal energy conversion and storage, and vehicle and aerospace operations (*3*–*10*). However, such control over heat is far behind the degree of control that has been achieved for electrons. Fundamentally, controlling thermal conductance using convenient handles—e.g., electric fields—has been a long-term but elusive dream for physicists and engineers, challenged by the spectral distribution of heat carriers and their complicated and weak interactions with external fields (*11*).

Advances in nanotechnology have greatly extended our understanding of and capabilities in the design and control of materials' properties and have provided opportunities for thermal control. For example, heat transport at low dimensions can be anomalous and can break Fourier's law. Early progress in demonstrating thermal rectification has been reported using carbon and boron nitride nanotubes with inhomogeneous mass loading (*12*). In such two-terminal devices, thermal conductance switching changes of up to 7% were observed for reversed heating directions. A great deal of theoretical and experimental work has been developed (*13*–*20*) to tune room-temperature thermal transport in solids relying mainly on structural phase changes or mass transfer, electrochemical scattering motion of ions, or other heat carriers. The usefulness of these approaches is largely limited not only by small switching ratios but also by intrinsically slow response speeds—typically from minutes to hours (i.e., usually far below 1 Hz). In this work, we demonstrate electrically gated three-terminal thermal switches that achieve high performance with >1-MHz switching speed, 1300% tunability in thermal conductance, and high reversibility over 1 million switching cycles. These values are substantial advancements over current solid-state thermal devices at room temperature.

Electrically gated three-terminal thermal devices

We illustrate the concept of our thermal switch with a symbolic diagram in Fig. 1A that is analogous to an electronic transistor. The device channel is contacted between a hot and cold thermal reservoir, with temperatures T_H and T_C, respectively. A third terminal using an electric field serves as the gate to control thermal conductance of the channel and thereby the heat flow. We designed this thermal switch and imaged the microstructure of our fabricated device [see supplementary materials (*21*) for details] with scanning electron microscopy (SEM) (Fig. 1B) from a tilted top-view angle and used high-resolution transmission electron microscopy (TEM) from a cross-sectional angle (Fig. 1C). We show a thin-film structure with an atomically resolved interface. This multilayer solid-state thin film consists of a self-assembled monolayer (SAM) of carboranethiol cage molecules on a bottom layer of gold (Au) from evaporative deposition and topped by single-layer graphene, which serves as the active conduction channel. Specifically, we achieved a cross-plane molecular junction using a SAM based on 9-SH-*o*-$C_2B_{10}H_{11}$ (O9) (*22*–*24*). These molecules form high-quality monolayers with minimum defect structures in complete monolayers limited to translational and rotational offsets in the registry of attachment to the substrate. We have shown that such layers provide the capability to tune the band alignment between contacts and the active layers in organic electronics (*24*). In this work, we show a typical scanning tunneling microscopy image of the carboranethiolate SAM and its closely packed and uniform structure with an atomically smooth exposed surface (Fig. 1D). The mechanism of the thermal switching is based on electrical control of atomic bond strengths for the molecular junctions that thereby controls thermal conductance.

We measured the thermal conductance (G) of the device per unit area, which shows a strong dependence of gate voltage (V_g) when applied across the molecular junction (Fig. 1E). When V_g is applied from 2.5 to −2.5 V, G exhibits a large monotonic increase from 10 to 134 MW/m^{2}K (red symbols). The thermal conductance control is >1300%, much higher than that of other demonstrations of electrical modulation (*15*, *16*). We also investigated molecular junctions where graphene was removed, and the G shows ambipolar behavior (blue symbols) that increases from 23 MW/m^{2}K when V_g = 2.5 V to 84 MW/m^{2}K when V_g = 0 V and then decreases to 8.0 MW/m^{2}K when V_g = −2.5 V. To test device reversibility, we performed the gate switching measurement for up to 1 million cycles. Our cycling measurements indicate that the gate switching on thermal conductance is highly reversible between the on and off states (Fig. 1F). This dynamic and reversible tunability of thermal conductance using an electric field effect demonstrates the robustness of these molecular thermal devices with high switching ratios (*15*, *16*).

First-principles theory

To elucidate the fundamental operation mechanisms, we performed first-principles calculations to determine the atomic charge distributions and thus the bonding energy (ε) of the molecular channel. We applied density functional theory (DFT) to calculate the spatial charge distribution over the atomic structure of Au-SAM-graphene (Fig. 2A) (*21*). Au and the S

[1]Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, CA 90095, USA. [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA. [3]California NanoSystems Institute, University of California, Los Angeles, CA 90095, USA. [4]Department of Bioengineering, University of California, Los Angeles, CA 90095, USA. [5]Department of Materials Science and Engineering, University of California, Los Angeles, CA 90095, USA.
*Corresponding author. Email: yhu@seas.ucla.edu
†These authors contributed equally to this work.

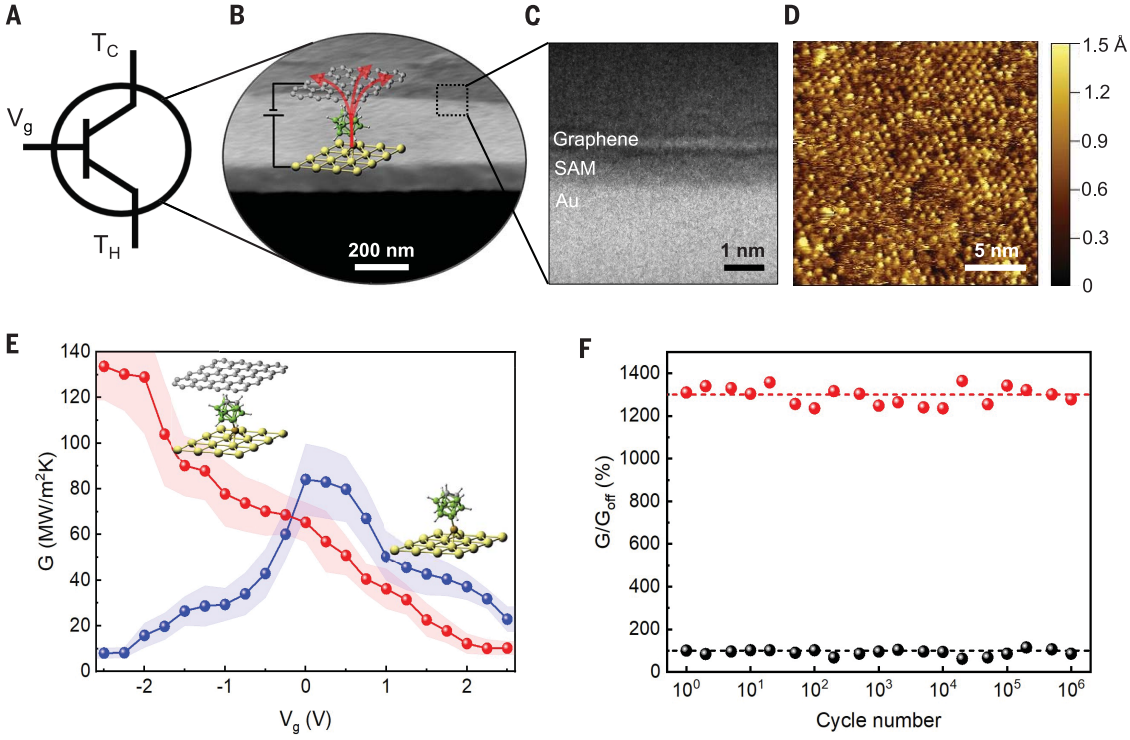


Fig. 1. Device design and transport performance of the molecular thermal switches. (**A**) Conceptual illustration of the three-terminal thermal device. (**B**) Tilted top-view SEM image of the microstructure, consisting of a monolayer of carborane-thiol cage molecules on a Au layer topped by single-layer graphene. (**C**) High-resolution cross-sectional TEM image of the device structure showing the atomically resolved interface. (**D**) Scanning tunneling microscopy image of the carborane-thiolate molecular monolayer on Au{111}. The image was acquired in constant-current mode at a −0.1-V bias under ambient conditions. (**E**) The measured thermal conductance (G) as a function of gate voltage (V_g). Red and blue curves represent device structures with and without graphene, respectively, as illustrated by the inset schematics. The shadowed background represents measurement uncertainty. (**F**) Reversibility test of electrical gating between ± 2.5 V on cycling measurements of the thermal device with graphene up to 1 million times. Red and black dashed guidelines indicate ratios of 1300% and 100%, respectively. The value of G_{off} is 10 MW/m^{2}K.

atoms (inside the SAM) form chemical bonds by sharing electrons, and the bond strength is characterized by the charge density difference (ρ_{diff}) (*25*)

$$\rho_{diff} = \rho_{Au+SAM} - (\rho_{Au} + \rho_{SAM}) \quad (1)$$

where ρ_{Au+SAM}, ρ_{Au}, and ρ_{SAM} are the charge density distributions of the bonded system, of the isolated Au surface, and of the isolated SAM, respectively. Thus, ρ_{diff} is responsible for the formation of the Au-SAM bond. To reveal the deterministic effects of V_g on electron sharing, we quantified the redistribution of electrons through the calculation of ρ_{diff} as a function of electric field.

We performed calculations for the Au-SAM system (i.e., without graphene). We illustrate the gating effects by plotting the variations of the charge density distributions ($\Delta\rho_{diff}$) under applied gate voltages (Fig. 2, B and C). The results indicate that the electrons at the bonding site—i.e., the Au-S bonds—have a large dynamic change. Electrons are pulled away from the S atom when a positive V_g is applied (Fig. 2B). By contrast, the electrons are more attracted toward the S atom on the carboranethiolate molecule when a negative V_g is applied (Fig. 2C). In either case, reduction in shared electrons between S and Au atoms eventually suppresses their covalent interactions and weakens their bond strength (*26*). Quantitatively, we calculated the Au-SAM bond energy (ε_{Au-SAM}) through first-principles theory (*21*) as a function of electrical voltage

$$\varepsilon_{Au-SAM} = (E_{Au} + E_{SAM}) - E_{Au+SAM} \quad (2)$$

where E_{Au+SAM}, E_{Au}, and E_{SAM} are the energy of the bonded system, that of the isolated Au, and that of the isolated SAM, respectively. Because the Au-SAM bond is one of the weakest covalent bonds in the system (*27*–*29*), our calculation shows that gate tuning of ε_{Au-SAM} is considerable, and the Au-S bond forms the thermal resistance inside the channel. Notably, ε_{Au-SAM} (blue solid curve) exhibits an ambipolar dependence as a function of voltage (Fig. 2D), similar to the measured trend for thermal conductance (Fig. 1E). Consistently, substantial changes in electrical conductance, hybridization, and presumably charge redistribution were observed at molecule-substrate interfaces as a result of motion (tilt) of adsorbed molecules (*30*).

We also performed first-principles calculations for a thermal device including the top graphene layer—i.e., the Au-SAM-graphene system. We applied nonlocal DFT to capture such long-range interactions accurately with an ab initio approach (*31*–*34*). Using Eq. 2, we determined the bond energy for SAM-graphene (Fig. 2D). We found a monotonic dependence for $\varepsilon_{SAM-graphene}$ (red curve) on voltage, which is a similar trend as observed in our measurements of gated thermal conductance. The calculation results are consistent with generally understood van der Waals interaction energies being derived from higher-order dipole-dipole interactions based on quantum perturbation theory (*35*). The bond energies increase with external electric fields because of the enhanced dipole moments determined by positive and negative charge separation. In addition, our theory shows strengthened Au-SAM bonding and its weak voltage dependence after adding a graphene layer (Fig. 2D, blue dashed line) because of the redistribution of charge density (*21*) at these interfaces (*36*). Therefore, our first-principles calculations indicate a fundamental tuning mechanism for thermal conductance by electric field–modulated charge redistribution. This family of carborane molecules are particularly well suited for this purpose, not only because of their robust self-assembly but because charge distribution within the cage is determined by the carborane cage selected as well as the point or points of attachment (*22*, *37*). We anticipate the potential for further enhancement of these devices by optimizing the polarizability and charge distributions of the molecular layer by selecting both the molecular

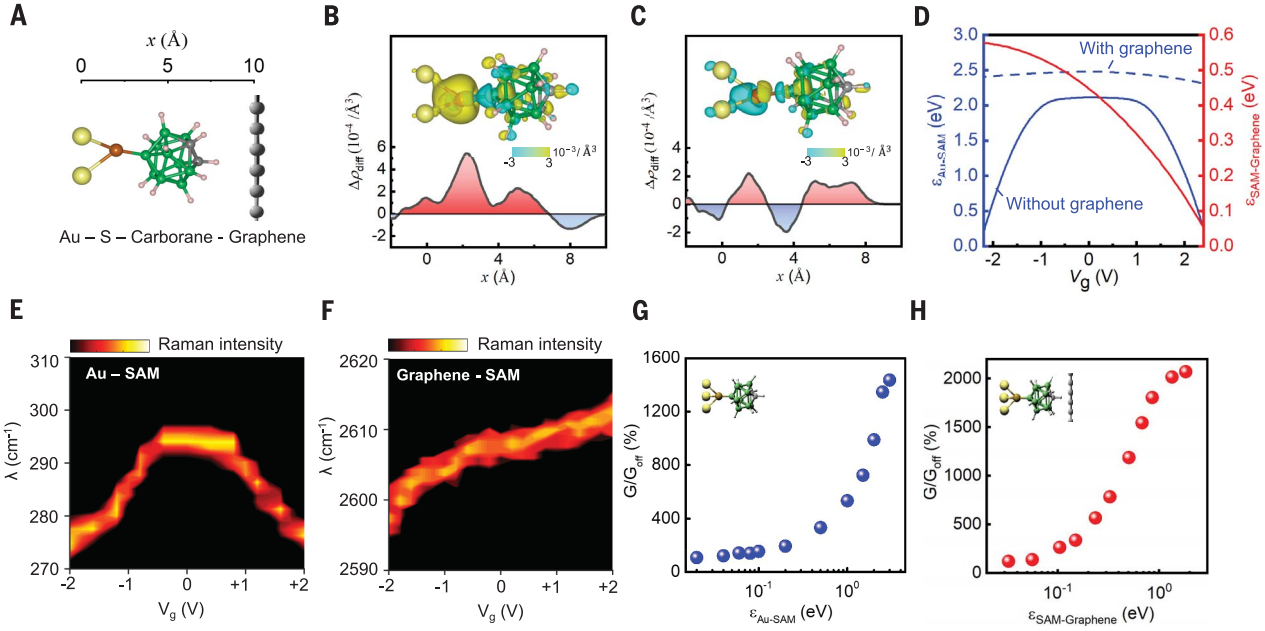


Fig. 2. Thermal switching operation mechanisms. (**A**) Atomic structure of the Au-monolayer(S-carborane)–graphene system. (**B** and **C**) Deviational charge density differences ($\Delta\rho_{diff}$) determined by DFT under an electric field of 0.6 V/Å (B) and −0.4 V/Å (C). Insets indicate the three-dimensional charge distributions. (**D**) Bond energies determined by DFT. Plotted as a function of V_g are the calculation results for the atomic bonds, including Au-S without graphene (blue solid line), Au-S with graphene (blue dashed line), and SAM-graphene (red line). (**E** and **F**) Experimental measurements of Raman peaks as a function of V_g. (E) Gate tuning of the stretching vibrational mode for Au-S. (F) Gate tuning of the 2D mode for graphene-SAM. (**G** and **H**) Molecular dynamics simulation results of thermal conductance as a function of bond energy for Au-SAM (G) and graphene-SAM (H) systems.

isomers of the cage molecules and the chemistries of attachment.

In situ vibrational spectroscopy

To test the calculated results, we experimentally measured the vibrational spectra corresponding to the interface bonding using in situ surface-enhanced Raman as a function of gate voltage. We applied the excitation laser with parallel incidence to the cross-plane of the thermal device film to study the interface vibrational modes. The A_1 and 2D Raman peaks are located at ~296 and 2608 cm^{-1}, respectively. The A_1 mode is the symmetric Au-S stretching vibration of Au-carboranethiolate (Fig. 2E) (*38*). As shown in the two-dimensional scanning image (Fig. 2E), with V_g sweeping from +2 to −2 V, the Raman peak first increases from 276.2 to 294.0 cm^{-1} and then decreases to 274.9 cm^{-1}. We observe ambipolar behavior consistent with the theoretically predicted bond energy variations (Fig. 2D) (*39*) and the observed ambipolar behavior of the measured conductance. For the Au-SAM-graphene system, we measured the evolution of the 2D mode Raman spectra as a function of V_g (Fig. 2F). The 2D Raman peak represents the double resonance process of photon scattering with two transverse optical (TO) phonons near the Brillouin edge and has a robust vibrational footprint frequency that quantifies the strain-dependent shift resulting from the spatial force heterogeneity. In this case, because of the curvature of the carborane molecules, the repulsive and attractive forces at the apex and shoulder of carborane stretch the graphene and induce lateral strains on the graphene through van der Waals interactions (*40*). Thus, stronger van der Waals interactions result in larger 2D Raman peak position shifts (*41*). The 2D Raman peak shows monotonic decreases from 2612.4 to 2599.1 cm^{-1} (a shift of 13.3 cm^{-1}) when the gate voltage is swept from +2 to −2 V (Fig. 2F), which indicates continuous and tunable modulation through van der Waals interactions under the applied electric field. The correlations between the van der Waals interactions and strain can be evaluated further from our measurements. The biaxial strain of graphene caused by the van der Waals interactions can be estimated from Fig. 2F to be ~0.08%, according to the 2D Raman peak shift coefficient of ~160 cm^{-1}/% versus biaxial strain in the linear elasticity regime (*41*). Because the strain is induced by the van der Waals interactions, the change of the interface bond energy can be estimated from the mechanical energy change. Considering a harmonic model for graphene with a stiffness of 150 N/m (*42*), the Raman peak shift estimates such energy changes as ~0.35 eV per molecule, which agrees with the predicted value from our DFT calculations (Fig. 2D). These in situ Raman measurements are consistent with electric field–induced changes in bond energy, predicted by our first-principles theory, as the dynamic tuning mechanism for the thermal switching behaviors. In addition, we simulated the thermal conductance to evaluate its dependence on bond energies using molecular dynamics (*43*). Our first principles–derived bond energies were input into the simulations to describe the interface interactions. The simulation domain mimics the experimental conditions without and with graphene (Fig. 2, G and H, respectively) under the varied bond energy for ε_{Au-SAM} and $\varepsilon_{SAM-graphene}$, which verifies the trend of gate tuning on thermal conductance.

High-speed thermal switching

As a further step, we performed direct measurements to characterize the thermal switching speeds. A key challenge for developing solid-state thermal switches is the intrinsically slow motion of heat and/or mass flow that usually limits thermal switching responses to minutes or hours. In this molecular thermal switch design, however, the driving mechanisms are field-induced, nearly instantaneous atomic-level charge and bonding dynamics, with the result that we achieved ultrahigh switching frequencies. We tested gating frequency using ultrafast optical spectroscopy based on the time-domain thermoreflectance (TDTR) technique. This technique is well suited for measuring the molecular switches because no physical contact is required with the sample, and the measurement provides

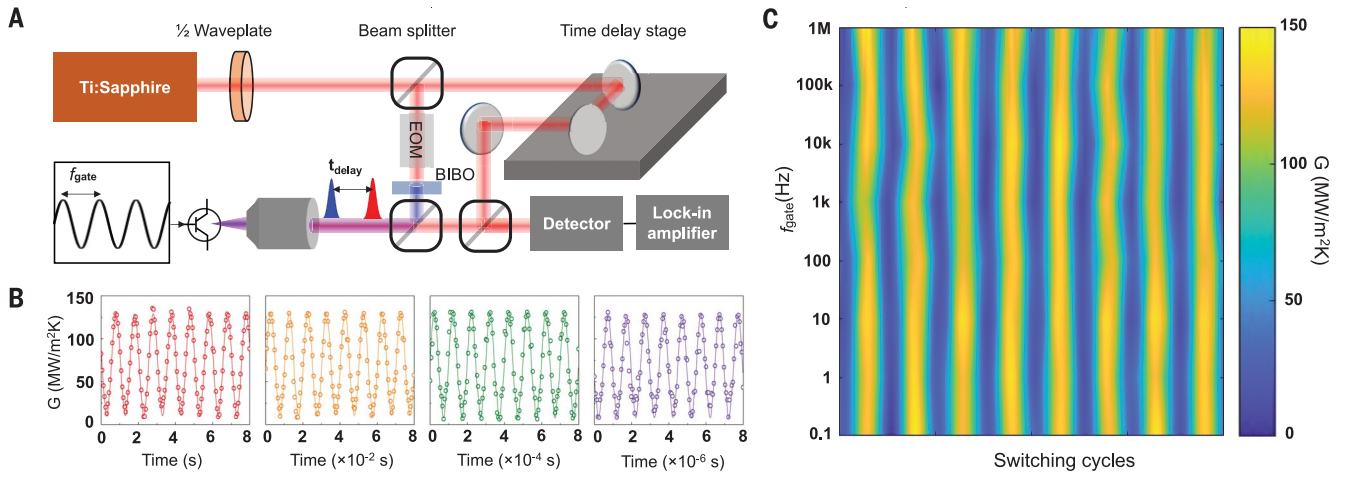


Fig. 3. High–switching speed measurements of electrically gated molecular thermal switch. (**A**) Schematic of experimental setup for time-dependent thermal conductance measurements under periodic gating voltage with varied gating frequency (f_{gate}). BIBO, bismuth triborate; EOM, electro-optic modulator; t_{delay}, time delay. (**B**) Experimental data of thermal conductance (circles) in response to gate fields with different frequencies (f_{gate}), plotted together with sine function waves of 1 Hz, 100 Hz, 10 kHz, and 1 MHz (solid lines). (**C**) Two-dimensional plot for thermal switching cycles versus scanning frequency response up to 1 MHz, measured at every decade. Note that 1 MHz is the instrument limit by the thermal detection bandwidth of the measurement setup.

high measurement resolution and accuracy. We have used TDTR to measure the thermal conductance of a wide range of materials, including molecular junctions (*4*, *9*, *17*, *43*–*47*). We used a train of 800-nm femtosecond laser pulses divided into a pump beam and a probe beam by a beam splitter (Fig. 3A). We used the pump beam to heat up the devices and the probe beam to detect the resulting temperature responses and determine the thermal conductance. The ultrafast measurements with subpicosecond time resolution enable effectively instantaneous thermal conductance determinations in real time upon gate tuning. To test the switching frequency (f_{gate}), time-dependent conductance changes versus periodically varied gating voltage are monitored. Our measured thermal conductance (circles) is reversibly switched between on and off states, following V_g with frequencies from 0.1 Hz to 1 MHz (Fig. 3B). The close synchronization between the thermal conductance and gating voltage (Fig. 3B) and the two-dimensional frequency scanning plot (Fig. 3C) indicates that the switching speed of our thermal switching is at least 1 MHz.

Conclusions

We have developed an electrically gated thermal switch based on solid-state molecular structures that demonstrates excellent performance at room temperature, with switching speeds above 1 MHz, switching ratios greater than 1300%, and high reversibility over 1 million gating cycles. We performed first-principles atomistic calculations and in situ vibrational spectra measurements to capture the dynamic physics of interface bonding and quantitatively explained our experimental results. The cage molecules that we used can be tuned to optimize polarizability and thus charge redistribution under gate control at the overlayer interfaces to enhance switching ratios further. Detailed calculations, as demonstrated above, guide both molecular and device design. With the availability of electrical control over heat flow and compatible fabrication with modern electronics, phonons can be integrated with electronics and photonics for future information processing. We anticipate that this concept will aid in the discovery of new opportunities in the control of thermal current for logic circuits as well as in thermal management and energy technologies.

REFERENCES AND NOTES

1. J. Bardeen, W. H. Brattain, *Phys. Rev.* **74**, 230–231 (1948).
2. The Nobel Prize in Physics 1956; https://www.nobelprize.org/prizes/physics/1956/summary/.
3. "Heating and Cooling" (US Department of Energy); https://www.energy.gov/energysaver/heating-and-cooling.
4. J. S. Kang, M. Li, H. Wu, H. Nguyen, Y. Hu, *Science* **361**, 575–578 (2018).
5. S. Chu, A. Majumdar, *Nature* **488**, 294–303 (2012).
6. M. M. Waldrop, *Nature* **530**, 144–147 (2016).
7. P. Ball, *Nature* **492**, 174–176 (2012).
8. International Technology Roadmap for Semiconductors, ITRS 2.0; www.itrs2.net/.
9. S. Li *et al.*, *Nature* **612**, 459–464 (2022).
10. Z. Qin, M. Li, J. Flohn, Y. Hu, *Chem. Commun.* **57**, 12236–12253 (2021).
11. J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (Oxford Univ. Press, 1960).
12. C. W. Chang, D. Okawa, A. Majumdar, A. Zettl, *Science* **314**, 1121–1124 (2006).
13. B. Li, L. Wang, G. Casati, *Appl. Phys. Lett.* **88**, 143501 (2006).
14. N. Li *et al.*, *Rev. Mod. Phys.* **84**, 1045–1066 (2012).
15. G. Wehmeyer, T. Yabuki, C. Monachon, J. Wu, C. Dames, *Appl. Phys. Rev.* **4**, 041304 (2017).
16. T. Swoboda, K. Klinar, A. S. Yalamarthy, A. Kitanovski, M. Munoz Rojo, *Adv. Electron. Mater.* **7**, 2000625 (2021).
17. J. S. Kang, M. Ke, Y. Hu, *Nano Lett.* **17**, 1431–1438 (2017).
18. G. Zhu *et al.*, *Nat. Commun.* **7**, 13211 (2016).
19. Q. Lu *et al.*, *Nat. Mater.* **19**, 655–662 (2020).
20. A. Sood *et al.*, *Nat. Commun.* **9**, 4510 (2018).
21. Materials and methods are available as supplementary materials.
22. J. N. Hohman *et al.*, *ACS Nano* **3**, 527–536 (2009).
23. J. C. Thomas *et al.*, *ACS Nano* **9**, 4734–4742 (2015).
24. J. Kim *et al.*, *Nano Lett.* **14**, 2946–2951 (2014).
25. L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (Cornell Univ. Press, 1960).
26. S. S. Shaik, P. C. Hiberty, *A Chemist's Guide to Valence Bond Theory* (Wiley, 2007).
27. W.-S. Liao *et al.*, *Science* **337**, 1517–1521 (2012).
28. A. M. Andrews, W. S. Liao, P. S. Weiss, *Acc. Chem. Res.* **49**, 1449–1457 (2016).
29. J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, G. M. Whitesides, *Chem. Rev.* **105**, 1103–1169 (2005).
30. Z. J. Donhauser *et al.*, *Science* **292**, 2303–2307 (2001).
31. H. Wu, H. Fan, Y. Hu, *Phys. Rev. B* **103**, L041203 (2021).
32. H. Fan, H. Wu, L. Lindsay, Y. Hu, *Phys. Rev. B* **100**, 085420 (2019).
33. M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **92**, 246401 (2004).
34. D. Chakraborty, K. Berland, T. Thonhauser, *J. Chem. Theory Comput.* **16**, 5893–5911 (2020).
35. M. Born, K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford Univ. Press, 1954).
36. Y. Cui, M. Li, Y. Hu, *J. Mater. Chem. C* **8**, 10568–10586 (2020).
37. J. C. Thomas *et al.*, *ACS Nano* **12**, 2211–2221 (2018).
38. A. Tlahuice-Flores, R. L. Whetten, M. Jose-Yacaman, *J. Phys. Chem. C* **117**, 12191–12198 (2013).
39. M. T. Dove, *Introduction to Lattice Dynamics* (Cambridge Univ. Press, 2005).
40. Y. Zhang *et al.*, *Nano Lett.* **18**, 2098–2104 (2018).
41. F. Ding *et al.*, *Nano Lett.* **10**, 3453–3458 (2010).
42. A. Zandiatashbar *et al.*, *Nat. Commun.* **5**, 3186 (2014).
43. M. Li *et al.*, *Adv. Mater.* **31**, e1901021 (2019).
44. J. S. Kang *et al.*, *Nat. Electron.* **4**, 416–423 (2021).
45. M. Li, J. S. Kang, Y. Hu, *Rev. Sci. Instrum.* **89**, 084901 (2018).
46. Y. Hu, L. Zeng, A. J. Minnich, M. S. Dresselhaus, G. Chen, *Nat. Nanotechnol.* **10**, 701–706 (2015).
47. J. S. Kang, H. Wu, Y. Hu, *Nano Lett.* **17**, 7507–7514 (2017).

ACKNOWLEDGMENTS

Funding: Y.H. acknowledges support from the National Institute of General Medical Sciences (NIGMS) under grant no. R35GM147391, the Alfred P. Sloan Research Foundation under grant no. FG-2019-11788, and the National Science Foundation (NSF) under award no. DMR-1753393. P.S.W. acknowledges support from the NSF under grant no. CHE-2004238. We acknowledge support from UCLA Nanolab and California NanoSystems Institute and the computational and storage services associated with the Hoffman 2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group; from the Bridges-2 at Pittsburgh

Supercomputing Center through allocation no. DMR180111 from Extreme Science and Engineering Discovery Environment supported by NSF grant no. 1548562; and from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program supported by NSF grants no. 2138259, no. 2138286, no. 2138307, no. 2137603, and no. 2138296. **Author contributions:** Y.H. designed the research; Y.H. and P.S.W. directed the study; M.L., Z.Q., and H.D.N. fabricated thermal devices and performed thermal and structural characterizations; H.W. and M.L. performed the ab initio calculations and molecular dynamics simulations; E.M.A., D.P.G., and M.L. prepared the molecules; and E.M.A., D.P.G., and T.L. collected the scanning tunneling microscopy data. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4297
Materials and Methods
Figs. S1 to S18
Table S1
References (*48*–*97*)

Submitted 13 February 2022; resubmitted 26 February 2023
Accepted 26 September 2023
10.1126/science.abo4297

RESTORATION ECOLOGY

Herbivory limits success of vegetation restoration globally

Changlin Xu[1], Brian R. Silliman[2], Jianshe Chen[1], Xincheng Li[1], Mads S. Thomsen[3,4], Qun Zhang[1], Juhyung Lee[5,6], Jonathan S. Lefcheck[7,8], Pedro Daleo[9], Brent B. Hughes[10], Holly P. Jones[11], Rong Wang[12], Shaopeng Wang[13], Carter S. Smith[2], Xinqiang Xi[14], Andrew H. Altieri[15], Johan van de Koppel[16,17], Todd M. Palmer[18], Lingli Liu[19], Jihua Wu[20], Bo Li[21], Qiang He[1]*

Restoring vegetation in degraded ecosystems is an increasingly common practice for promoting biodiversity and ecological function, but successful implementation is hampered by an incomplete understanding of the processes that limit restoration success. By synthesizing terrestrial and aquatic studies globally (2594 experimental tests from 610 articles), we reveal substantial herbivore control of vegetation under restoration. Herbivores at restoration sites reduced vegetation abundance more strongly (by 89%, on average) than those at relatively undegraded sites and suppressed, rather than fostered, plant diversity. These effects were particularly pronounced in regions with higher temperatures and lower precipitation. Excluding targeted herbivores temporarily or introducing their predators improved restoration by magnitudes similar to or greater than those achieved by managing plant competition or facilitation. Thus, managing herbivory is a promising strategy for enhancing vegetation restoration efforts.

Vegetation is a primary foundation of Earth's ecosystems, spanning from rainforests and grasslands on land to seagrass beds and kelp forests in coastal oceans. Vegetation in many ecosystems, however, has been extensively degraded, resulting in the loss of biodiversity and critical services to humanity (*1*, *2*). The restoration of degraded vegetation has been increasingly adopted to recover ecosystem functions and services (e.g., carbon sequestration and flood mitigation), which contributes to efforts to achieve many of the United Nations (UN) Sustainable Development Goals (*3*, *4*). To fulfill commitments in the UN Decade on Ecosystem Restoration, governments, nongovernmental organizations, and local communities are seeking effective strategies to restore vegetation.

Restoration of vegetated ecosystems has traditionally adopted a bottom-up framework (i.e., the "Field of Dreams" hypothesis), which posits that plants (and subsequently higher trophic levels) will recover through natural plant propagation if the cause of degradation (e.g., tree logging, farming, or plant invasion) is removed, sometimes with additional activities to ameliorate changes to the physical environment (e.g., hydrological modification or soil amendment) (*5*). We refer to this restoration approach as natural regeneration (also known as passive restoration) (*5*). A more active approach is planting seeds, seedlings, or other plant propagules to expedite vegetation recovery. We refer to this approach as planted restoration (also known in some literature as active restoration) (*6*, *7*).

Plant abundance and diversity in ecosystems that are restored by these bottom-up approaches often do not fully recover, even after decades (but there are exceptions, especially at tropical sites) (*7*, *8*). This limited recovery may be caused by unrecognized top-down control of vegetation by herbivores or indirectly by the loss of predators that control herbivore populations (*9*, *10*), which limits vegetation establishment or growth at restoration sites. Top-down control has been documented extensively as a key force that structures vegetation in relatively undegraded ecosystems where trophic interactions among plants, herbivores, and predators have stabilized (*11*, *12*). Despite studies on specific ecosystems or consumers (*13*, *14*), there has not yet been a global assessment of top-down effects on restoration. This gap hampers progress in upscaling restoration and leaves the recent initiative of trophic rewilding, in which reintroducing predators is recommended to promote biodiverse and self-regulating ecosystems (*15*), unsubstantiated by broad empirical evidence.

We present an assessment of top-down control of vegetation under restoration globally by assembling a dataset of consumer effects on vegetation, with 1898 field experimental tests conducted in 64 countries and published in 451 articles [the Global Consumer Effects dataset (*16*); see figs. S1 and S2, table S1, and materials and methods (*17*)]. This dataset includes tests of consumer effects on vegetation by herbivore exclusion or addition (1883 tests) or predator reintroduction (15 tests) at relatively undegraded (550 tests), natural regeneration (1049 tests), and planted restoration (299 tests) sites. We considered relatively undegraded sites (hereafter, undegraded sites) to be those not impaired by recent anthropogenic disturbance, although we recognize that few sites on Earth remain completely undisturbed. These tests encompassed terrestrial and aquatic (freshwater and marine) biomes across all six vegetated continents (Fig. 1A) and spanned tropical, subtropical, and temperate regions (274, 773, and 851 tests, respectively, with tropical aquatic studies considerably less represented; Fig. 1C) and broad gradients of temperature and precipitation (Fig. 1B). Each test included measures of vegetation abundance (density, percent cover, biomass, or survival; 1434 tests) or diversity (species richness, Shannon diversity, Simpson's evenness, or Pielou's evenness; 464 tests), both in control (X_C) and in herbivore exclusion or predator reintroduction (X_T) treatments [the relatively small number of herbivore addition studies were omitted in our main analyses (*17*); fig. S3]. These vegetation measures represent proxies for desirable restoration outcomes because they correlate with valued ecosystem functions such as carbon sequestration and ecosystem stability (*18*, *19*). For each test, we calculated the effect size of herbivores on vegetation abundance and diversity using the log response ratio, $\ln(X_C/X_T)$, with negative and

[1]MOE Key Laboratory for Biodiversity Science and Ecological Engineering, National Observations and Research Station for Wetland Ecosystems of the Yangtze Estuary, School of Life Sciences, Fudan University, Shanghai, China. [2]Nicholas School of the Environment, Duke University, Beaufort, NC, USA. [3]Marine Ecology Research Group and Centre for Integrative Ecology, School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. [4]Department of Bioscience, Aarhus University, Roskilde, Denmark. [5]Marine Science Center, Northeastern University, Nahant, MA, USA. [6]Department of Oceanography and Marine Research Institute, Pusan National University, Busan, Republic of Korea. [7]Tennenbaum Marine Observatories Network and MarineGEO Program, Smithsonian Environmental Research Center, Edgewater, MD, USA. [8]University of Maryland Center for Environmental Science, Cambridge, MD, USA. [9]Instituto de Investigaciones Marinas y Costeras (IIMyC), UNMdP, CONICETC, Mar del Plata, Argentina. [10]Department of Biology, Sonoma State University, Rohnert Park, CA, USA. [11]Department of Biological Sciences and Institute for the Study of the Environment, Sustainability, and Energy, Northern Illinois University, DeKalb, IL, USA. [12]School of Ecological and Environmental Sciences, Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, East China Normal University, Shanghai, China. [13]Institute of Ecology, College of Urban and Environmental Sciences, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing, China. [14]Department of Ecology, School of Life Science, Nanjing University, Nanjing, Jiangsu, China. [15]Department of Environmental Engineering Sciences, University of Florida, Gainesville, FL, USA. [16]Department of Estuarine and Delta Systems, Royal Netherlands Institute for Sea Research, Yerseke, Netherlands. [17]Conservation Ecology Group, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands. [18]Department of Biology, University of Florida, Gainesville, FL, USA. [19]State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China. [20]State Key Laboratory of Herbage Improvement and Grassland Agro-Ecosystems, and College of Ecology, Lanzhou University, Lanzhou, Gansu, China. [21]Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology and Centre for Invasion Biology, Institute of Biodiversity, School of Ecology and Environmental Science, Yunnan University, Kunming, Yunnan, China.
*Corresponding author. Email: he_qiang@hotmail.com

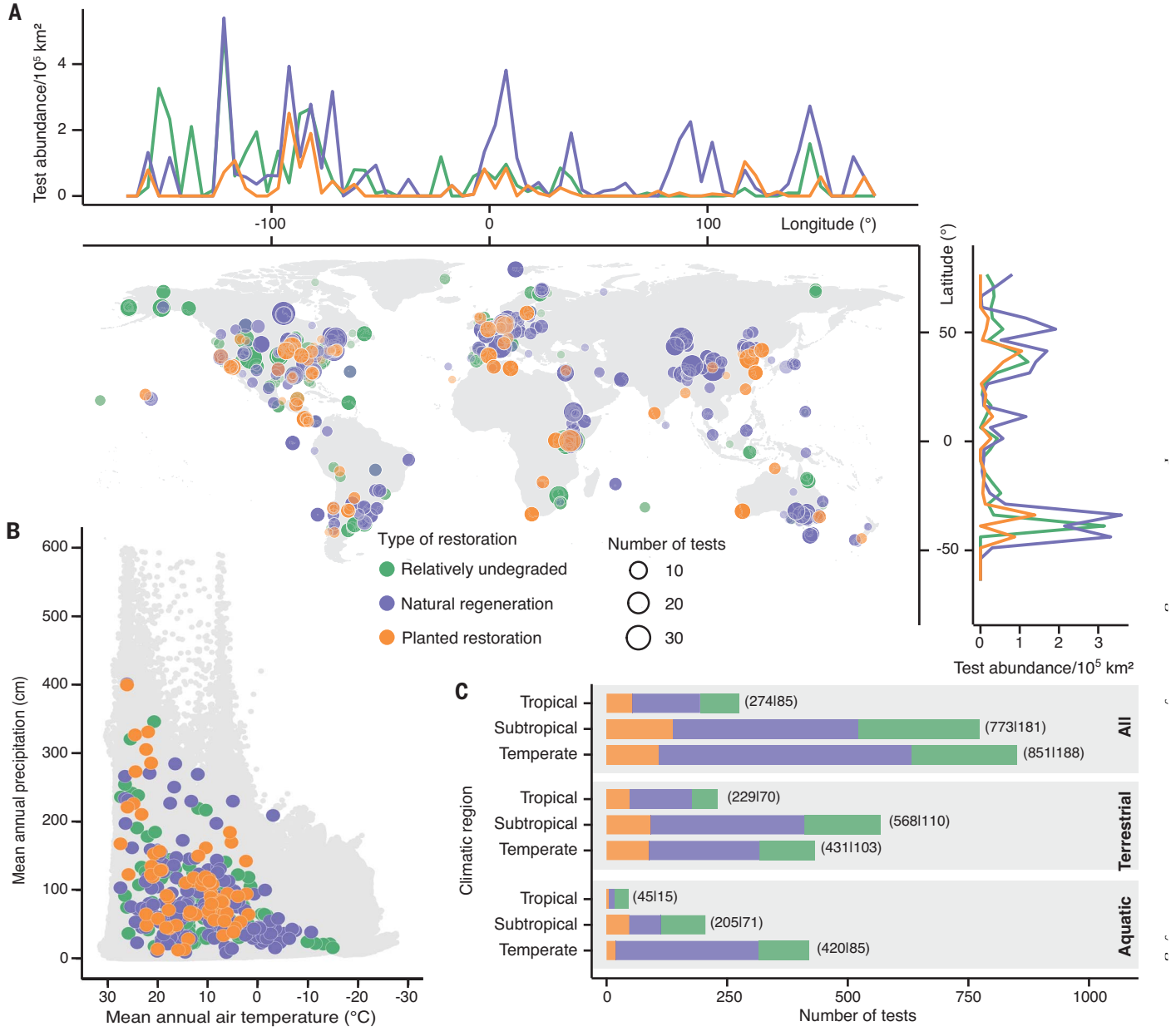


Fig. 1. Global patterns in the study of consumer effects on vegetation. (**A**) Global distribution of tests at natural regeneration, planted restoration, and relatively undegraded sites. The insets show the distributions of the abundance of tests per 10^5 km^2 of land area (excluding Antarctica) along latitudinal and longitudinal gradients (one point of test abundance, 41, between longitude −165° and −170° was omitted from the figure for visual clarity). Tests on aquatic vegetation that occur primarily along shorelines were also included to compute test abundance per unit land area. The equidistant cylindrical projection, instead of an equal-area projection, is used for the background map to align it with the insets that show test abundance. (**B**) Distribution of tests in the global climate space defined by MAT and MAP (gray points represent 1,000,000 randomly sampled points of global land areas, and colored points represent tests included in our dataset; 88 of the 574 coastal marine tests that did not have standardized climate data were omitted from the figure). (**C**) Number of tests in different biomes (terrestrial and aquatic) and climatic regions (tropical, subtropical, and temperate). Sample sizes are shown as number of tests|number of articles to the right of each bar.

positive values indicating negative and positive effects of herbivores, respectively. For each test, we also collected several covariates to identify moderators of variation in effect sizes (*17*): biotic (e.g., plant life forms such as macroalgal, herbaceous, or woody; herbivore size classes such as invertebrate, small vertebrate, and large vertebrate; the provenance of plants and herbivores), climatic [mean annual air temperature (MAT) and mean annual precipitation (MAP)], and methodological (e.g., study duration and plot size, that is, the area where herbivores were removed).

Additionally, to assess the effect size of managing herbivory (through herbivore exclusion or predator reintroduction) versus managing other widely considered biotic drivers of vegetation restoration, we assembled a global dataset of 696 tests on the effects of managing plant competition and facilitation from 167 articles [the Global Plant Interactions and Restoration dataset (*16*); fig. S1 and table S1] (*17*). Each of these tests measured vegetation abundance in control (X_C) and in competitor removal or facilitator inclusion (X_T) treatments at a natural regeneration or planted restoration site. Effect sizes of competitor removal or

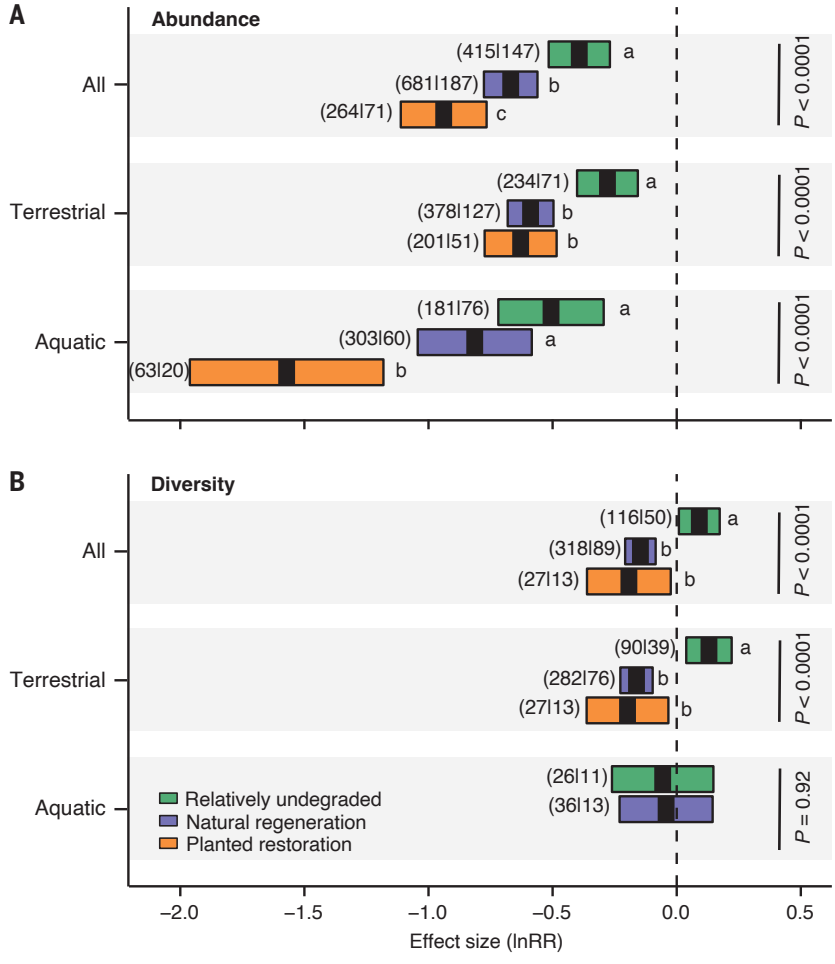


Fig. 2. Herbivores exert stronger effects on vegetation at restoration sites than at undegraded sites globally. (**A** and **B**) Herbivore effect sizes on plant abundance (A) and diversity (B) at natural regeneration, planted restoration, and relatively undegraded sites. Analyses were conducted for studies in terrestrial, aquatic, and all (combined) biomes. Effect sizes are log response ratios (lnRR), with negative and positive values indicating negative and positive effects of herbivores, respectively. Center bolded lines represent mean effect sizes, and error bars represent 95% CI. Sample sizes are shown as number of tests|number of articles to the left of each bar. Statistical significances are shown with *P* values. Within each analysis (shaded area), bars that share a letter (i.e., a, b, or c) do not differ significantly from one another ($P > 0.05$, based on post hoc multiple comparisons with Holm correction). Relatively undegraded, natural regeneration, and planted restoration sites are indicated with green, purple, and orange boxes, respectively.

facilitator inclusion were calculated as $\ln(X_T/X_C)$, with positive values indicating positive effects of management. We then compared these effect sizes with those of herbivore exclusion or predator reintroduction calculated for tests in the Global Consumer Effects dataset. Our analyses are robust to inclusion of tests with zero values (table S2) and potential publication bias (fig. S4 and tables S3 to S6).

Global, amplified effects of herbivores on vegetation under restoration

We first assessed the mean effect sizes [with 95% confidence interval (CI)] of herbivores at restoration (natural regeneration and planted restoration) sites in comparison to undegraded sites. We found strong negative effects of herbivores on the abundance of vegetation under restoration across terrestrial and aquatic biomes globally. Although herbivores naturally reduced vegetation abundance at undegraded sites (−32% on average), reductions were significantly stronger at restoration sites (−52%; Fig. 2A). This finding was corroborated by supplementary analyses that (i) focused on studies that tested herbivore effects at both undegraded and natural regeneration sites (fig. S5), (ii) used more-specific vegetation performance measures (fig. S6), (iii) accounted for differences in sample size through bootstrap resampling (table S7) (*17*), and (iv) accounted for potential effects of covariates through multivariate modeling (table S8) (*17*). Subsequently, vegetation recovery was often slower and less complete if herbivores were unchecked at restoration sites (fig. S7). Indeed, herbivore density was generally higher at restoration sites than at undegraded sites, and greater densities of herbivores often more strongly suppressed vegetation abundance (fig. S8). Herbivores (generalists in particular) can retreat during disturbance and rapidly reestablish (*20*), which likely contributes to amplified herbivore effects at restoration sites. Although herbivores might suppress vegetation abundance more strongly at high productivity sites (fig. S9), plant productivity was often lower at restoration sites than at undegraded sites (fig. S9) and thus could not explain the amplified herbivore effects at restoration sites.

We also examined whether the amplified effects of herbivores on vegetation abundance at restoration sites were consistent among different biomes, climatic regions, and types of restoration. We found amplified herbivore effects in both terrestrial and aquatic biomes (Fig. 2A) for both natural regeneration and planted restoration, which used different types of propagules like seeds, seedlings, or adults (Fig. 2A and table S9), and in tropical and subtropical regions (fig. S10). In temperate regions, however, herbivore effects were strongly negative at both undegraded and restoration sites (fig. S10). Our results also revealed some fundamental differences in herbivore effects between terrestrial and aquatic biomes. Herbivore effects on vegetation abundance were similar between natural regeneration and planted restoration sites in terrestrial biomes (grasslands and forests), whereas in aquatic biomes (freshwater wetlands and marine vascular plant systems; table S10), herbivore effects were much stronger at planted restoration sites than at natural regeneration sites (Fig. 2A and fig. S11). This pattern, however, held only when natural regeneration sites had target plants already present at the beginning of the study (fig. S12). This is likely in part because aquatic herbivores are more generalized in habitat use relative to their terrestrial counterparts (*21*) and may tend to concentrate on immature, regenerating plants rather than mature, defended plants.

Herbivores have generally been shown to increase plant diversity in naturally productive ecosystems by reducing the dominance of competitively superior plant species (*22*). Our global synthesis corroborates this paradigm at undegraded sites but also reveals that herbivores generally reduced plant diversity at restoration sites (Fig. 2B and fig. S13) (*17*). This reversal was found mainly in terrestrial biomes (from 14% at undegraded sites to −15% at restoration sites; Fig. 2B), which included grasslands and forests (fig. S11) across different climatic regions (fig. S10). Supplementary analyses showed that plant productivity was lower at restoration sites than at undegraded sites (fig. S9) and that herbivores more strongly

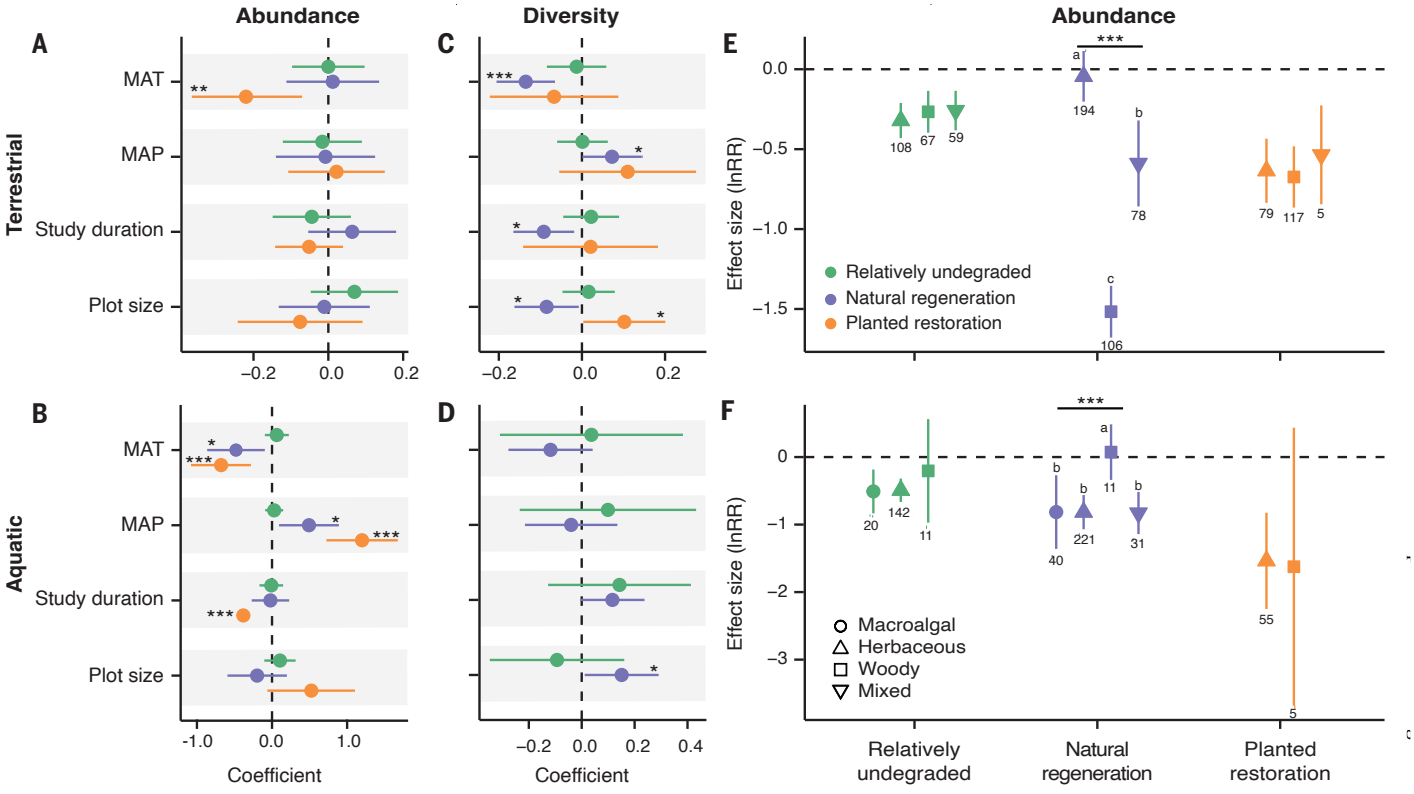


Fig. 3. Herbivore effects are contingent on climatic and biotic factors. (**A** and **B**) Coefficients of continuous covariates for the effect size of herbivores on vegetation abundance, estimated using a global model that included all covariates considered (*17*). (**C** and **D**) Coefficients of continuous covariates estimated using mixed-effects models on plant diversity (no studies were available for planted restoration in aquatic biomes). In (A) to (D), all coefficients were estimated using covariates that were standardized to have a mean of zero and a standard deviation of one; negative and positive coefficients indicate that the negative effects of herbivores strengthen and weaken with increases in a covariate, respectively (see fig. S15 for changes in the effect sizes of herbivores with covariates in original, untransformed units). (**E** and **F**) Mean effect sizes of herbivores on the abundance of different plant life forms (the number of tests is shown below each bar; see figs. S16 to S18 for additional categorical covariates). Effect sizes are log response ratios (lnRR), and negative values correspond to negative effects of herbivores on vegetation abundance. Panels (A), (C), and (E) show data for terrestrial biomes, and panels (B), (D), and (F) show data for aquatic biomes. In all panels, data are shown as means ± 95% CI. Statistical significances are indicated: ***$P < 0.001$, **$P < 0.01$, and *$P < 0.05$. In (E) and (F), bars that share a letter (i.e., a or b) do not differ significantly from one another ($P > 0.05$, based on post hoc multiple comparisons with Holm correction). Groups with <5 tests (which require further research) were omitted from the figures and discussions throughout this paper. In all panels, natural regeneration, planted restoration, and relatively undegraded sites are indicated by purple, orange, and green symbols, respectively.

suppressed plant diversity where plant productivity was lower (fig. S14), possibly owing to lower productivity sites having smaller plant populations that are more easily eliminated. This explanation for the reversal is supported by multiple studies that show that herbivores increase plant diversity at high-productivity sites but decrease it at low-productivity sites (*22*, *23*). Additionally, degraded ecosystems were often characterized by a higher density of generalist herbivores that are more likely to suppress plant diversity by removing early successional plant species (fig. S8) (*24*), which invest more in tolerance to environmental stressors than in defenses against herbivory (*25*). However, no such reversal was detected in aquatic biomes (Fig. 2B), where plant diversity is generally low and restoration efforts often targeted a single foundation species (*26*). We found no significant effect of herbivores on the recovery rate and completeness of plant diversity, although studies remained relatively few (fig. S7) or often lasted only a few years (table S11).

Climatic conditions are among the key moderators of herbivore effects at restoration sites

Exploring variation in the effect sizes of herbivores (*17*), we found that moderators of herbivore effects often differed between restoration and undegraded sites. First, we found climatic conditions, including MAT and MAP, to be among the key moderators at restoration sites but not at undegraded sites (Fig. 3, A to D; fig. S15; and tables S12 and S13). In both terrestrial and aquatic biomes, herbivores often more strongly suppressed vegetation abundance at sites with higher MAT (but this pattern was absent at terrestrial natural regeneration sites; Fig. 3). This finding corroborates the view that high temperatures generally increase herbivory (*27*), although this effect of temperature may be amplified or offset by its effect on plant growth, which varies depending on habitat moisture (*28*). In terrestrial biomes, herbivore effects on vegetation abundance did not vary significantly with MAP, whereas in aquatic biomes, effects were more negative with lower MAP (Fig. 3), perhaps because aquatic vegetation is more stressed from periodic drying and higher densities of herbivores during dry periods (*29*). Climate also explained variation in herbivore effects on plant diversity at natural regeneration sites, with herbivores exerting stronger negative effects in regions with high MAT and low MAP, a pattern that is significant in terrestrial biomes but not in low-diversity aquatic biomes (Fig. 3). In contrast to restoration sites, these climatic signals were nonsignificant at undegraded sites (Fig. 3, A to D). This difference suggests that systems under restoration may be more vulnerable to herbivory in warmer, drier climates.

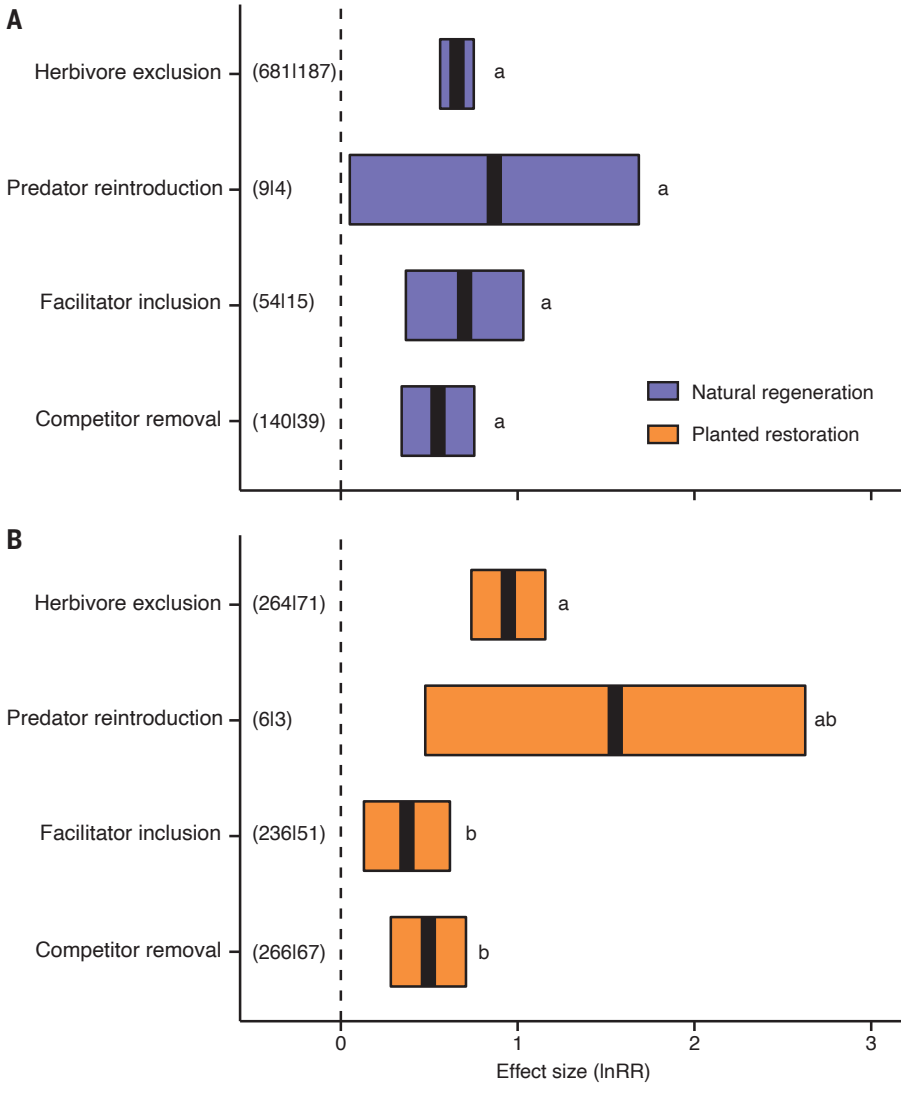


Fig. 4. Effects of managing herbivory in comparison to managing plant-plant interactions. Management of herbivory occurred by herbivore exclusion or predator reintroduction, whereas management of plant-plant interactions occurred by facilitator inclusion or competitor removal. (**A** and **B**) Effect sizes of management for natural regeneration (A) and planted restoration (B). Effect sizes are log response ratios (lnRR), and positive values indicate that management enhanced vegetation abundance. Center bolded lines represent mean effect sizes, and error bars represent 95% CI. Sample sizes are shown as number of tests|number of articles to the left of each bar. Bars that share a letter (i.e., a and/or b) do not differ significantly from one another ($P > 0.05$, based on post hoc multiple comparisons with Holm correction).

The effect sizes of herbivores at restoration sites also varied across plant functional groups. Herbivores more strongly reduced the abundance of woody versus herbaceous and mixed vegetation under natural regeneration in terrestrial biomes (Fig. 3E), where early successional shrubs and trees often have relatively few defenses against herbivory (*25*). This finding corroborates studies that have shown that herbivore exclusion accelerates woody encroachment in grasslands (*30*) and might decrease plant functional diversity in tropical forests (*31*). In aquatic biomes, however, herbivore effects were stronger for herbaceous and macroalgal vegetation than for woody vegetation under natural regeneration (Fig. 3F), likely because woody plants in aquatic biomes are less palatable or better defended against herbivory (*32*). Nonetheless, these patterns disappeared with planted restoration, where herbivore effects were generally strong regardless of biome and plant life form (Fig. 3, E and F). This finding suggests that the susceptibility of propagules used in planted restoration may not mirror that of naturally regenerating plants (*33*). Furthermore, herbivores often reduced the abundance of naturally regenerating native plants more than exotic plants (fig. S16), a pattern consistent with the enemy release hypothesis (*34*).

Not surprisingly, variations in the effect size of herbivores on vegetation abundance and diversity were additionally explained by herbivore functional groups, the cause of degradation, and the study methodology, including plot size and study duration (Fig. 3 and figs. S15, S17, and S18). For example, herbivore effects might strengthen with study duration (Fig. 3, A to D). This supports the idea that herbivores have accumulating effects across plant generations or successional stages (*25*), although the regeneration of woody plants can take much longer than the time period covered in most studies (table S11). Herbivore effects on plant diversity might strengthen or weaken with plot size, whereas effects on plant abundance at restoration sites were generally observed regardless of plot size (Fig. 3, A to D). Although aquatic restoration studies often used small plots (table S11), many terrestrial natural regeneration studies with >1-ha plots consistently reported strong herbivore effects on vegetation abundance (mean effect size: −0.73; 95% CI: −1.00 to −0.46) and diversity (mean effect size: −0.31; 95% CI: −0.54 to −0.09). These findings, along with well-documented large-scale impacts of changes in herbivore populations on vegetation (*9*, *35*), suggest that the effects of herbivores may scale up to real-world restoration settings.

Managing herbivory to enhance restoration success

The effects of herbivores at restoration sites can be managed through plant-based interventions, such as prioritizing revegetation at sites less affected by herbivores or using planting strategies that are robust to herbivory (e.g., planting at high densities or with grazing-resistant species). Our synthesis found that excluding herbivores with physical exclosures, insecticides, or deterrents, which is a consumer-based intervention more suited for small-scale restoration, increased vegetation abundance by 93 and 158% at natural regeneration and planted restoration sites, respectively, relative to plots open to herbivores (Fig. 4). Introducing predators, another consumer-based intervention, increased vegetation abundance by 138 and 372% at natural regeneration and planted restoration sites, respectively (variations in these effects were likely inflated by small sample size; Fig. 4), which supports calls for trophic rewilding (*15*). Both types of consumer-based intervention generally improved restoration by magnitudes that were comparable to or greater than those achieved by managing plant competition or facilitation (Fig. 4) across terrestrial and aquatic biomes (table S14) in tropical, subtropical, and temperate regions (fig. S19), including when accounting for sample-size differences through bootstrap resampling (table S15) (*17*).

Our findings offer insights toward achieving myriad restoration commitments. Massive

revegetation efforts are being implemented globally, including the Bonn Challenge (*36*), Africa's Great Green Wall (*37*), and the Blue Carbon Initiative (*38*). By demonstrating global, substantial impacts of herbivores on the abundance and diversity of vegetation under restoration, our study suggests that revegetation efforts, if implemented merely by removing the cause of degradation, recreating abiotic conditions, or planting propagules, are unlikely to achieve maximal outcomes. Rather, substantial improvements can be achieved by comanaging herbivory (by either plant- or consumer-based approaches). By revealing climates and other moderators of variations in herbivore effects at restoration sites, our study can help restoration practitioners pinpoint where and when managing herbivory may be particularly crucial, including in the tropics (Fig. 1C), where global priority areas for vegetation restoration are concentrated (*39*), and in hot, dry regions as well as in the years ahead with respect to future climates. Indeed, as climate change and human activities, which are often beyond the immediate control of local managers, continue to disrupt food webs and affect vegetation through top-down processes (*10*, *40*, *41*), managing herbivory may become increasingly relevant and tractable for enhancing recovery and resilience (*42*).

Interpretation of our results within the following context can provide guidelines for restoration practice. First, the causes of amplified herbivore effects are often site-specific, including (i) predator loss (*9*), (ii) increases in herbivore density or consumption (*43*), and/or (iii) decreases in plant resource and resistance (*40*), each of which requires distinct interventions. For certain herbivore species that are endangered or threatened or that may facilitate plant establishment through ecosystem engineering, by preferentially consuming competitive weeds or by promoting seed dispersal at low population densities (*44*, *45*), plant-based interventions may be more appropriate than consumer-based interventions, so that vegetation restoration serves as a foundation for restoring the whole ecosystem, including endangered or threatened fauna. Second, restoration practices should be bounded within the broader socioeconomic context. Reintroducing wild predators, for example, may be undesirable because of human-wildlife conflicts in farmed landscapes or urban environments. In such cases, managing herbivory through plant-based interventions or by introducing predator cues (mimics, sounds, or chemical signatures) may be possible solutions, especially where predator effects are predominantly driven by fear (*46*). Third, although we focused on herbivores (*17*), best practices should combine approaches and account for other biotic (e.g., parasites, omnivores, and microbes) and abiotic (e.g., nutrient and hydrology) factors (*11*, *47*), including potential synergistic or additive interactions of herbivores with abiotic stressors at restoration sites (*47*). Developing more-integrative restoration approaches is critical for fulfilling vegetation's enormous capacity for mitigating climate change, supporting biodiversity, and promoting sustainable development.

REFERENCES AND NOTES

1. J. T. Overpeck, D. D. Breshears, *Science* **372**, 786–787 (2021).
2. P. I. Macreadie *et al*., *Nat. Rev. Earth Environ.* **2**, 826–839 (2021).
3. International Resource Panel (IRP), "Land restoration for achieving the Sustainable Development Goals: An international resource panel think piece" (United Nations Environment Programme, 2019).
4. R. K. F. Unsworth, L. C. Cullen-Unsworth, B. L. H. Jones, R. J. Lilley, *Science* **377**, 609–613 (2022).
5. M. A. Palmer, R. F. Ambrose, N. L. Poff, *Restor. Ecol.* **5**, 291–300 (1997).
6. N. Shackelford *et al*., *Nat. Ecol. Evol.* **5**, 1283–1290 (2021).
7. H. P. Jones *et al*., *Proc. Biol. Sci.* **285**, 20172577 (2018).
8. D. Moreno-Mateos *et al*., *Nat. Commun.* **8**, 14163 (2017).
9. J. A. Estes *et al*., *Science* **333**, 301–306 (2011).
10. R. Dirzo *et al*., *Science* **345**, 401–406 (2014).
11. Q. He, B. R. Silliman, *Ecol. Monogr.* **86**, 278–294 (2016).
12. S. Jia *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6237–6242 (2018).
13. J. Watson, J. A. Estes, *Ecol. Monogr.* **81**, 215–239 (2011).
14. A. Filazzola *et al*., *Ecol. Lett.* **23**, 1298–1309 (2020).
15. J. C. Svenning *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 898–906 (2016).
16. C. Xu *et al*., Data from: Herbivory limits success of vegetation restoration globally. Zenodo (2023); https://doi.org/10.5281/zenodo.8338052.
17. Materials and methods are available as supplementary materials.
18. Y. Feng *et al*., *Science* **376**, 865–868 (2022).
19. F. Isbell *et al*., *Nature* **477**, 199–202 (2011).
20. T. D. Schowalter, *Annu. Rev. Entomol.* **57**, 1–20 (2012).
21. M. E. Hay, *Trends Ecol. Evol.* **6**, 362–365 (1991).
22. H. Olff, M. E. Ritchie, *Trends Ecol. Evol.* **13**, 261–265 (1998).
23. N. A. McMillan, K. E. Kunkel, D. L. Hagan, D. S. Jachowski, *Restor. Ecol.* **27**, 379–388 (2019).
24. M. Marvier, P. Kareiva, M. G. Neubert, *Risk Anal.* **24**, 869–878 (2004).
25. D. W. Davidson, *Oikos* **68**, 23–35 (1993).
26. B. R. Silliman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14295–14300 (2015).
27. E. Hamann, C. Blevins, S. J. Franks, M. I. Jameel, J. T. Anderson, *New Phytol.* **229**, 1894–1910 (2021).
28. G. Yvon-Durocher, A. P. Allen, J. M. Montoya, M. Trimmer, G. Woodward, *Adv. Ecol. Res.* **43**, 267–313 (2010).
29. B. R. Silliman, J. van de Koppel, M. D. Bertness, L. E. Stanton, I. A. Mendelssohn, *Science* **310**, 1803–1806 (2005).
30. Z. S. Venter, M. D. Cramer, H. J. Hawkins, *Nat. Commun.* **9**, 2272 (2018).
31. Y. Souza, N. Villar, V. Zipparro, S. Nazareth, M. Galetti, *J. Ecol.* **110**, 845–859 (2022).
32. J. D. Parker, C. C. Caudill, M. E. Hay, *Oecologia* **151**, 616–625 (2007).
33. M. Conrady *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2219664120 (2023).
34. R. M. Keane, M. J. Crawley, *Trends Ecol. Evol.* **17**, 164–170 (2002).
35. G. P. Asner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4947–4952 (2009).
36. M. Verdone, A. Seidl, *Restor. Ecol.* **25**, 903–911 (2017).
37. A. Mirzabaev, M. Sacande, F. Motlagh, A. Shyrokaya, A. Martucci, *Nat. Sustain.* **5**, 17–25 (2022).
38. D. Herr, E. Landis, "Coastal blue carbon ecosystems: Opportunities for nationally determined contributions. Policy brief" (International Union for Conservation of Nature and The Nature Conservancy, 2016).
39. B. B. N. Strassburg *et al*., *Nature* **586**, 724–729 (2020).
40. J. M. Tylianakis, R. K. Didham, J. Bascompte, D. A. Wardle, *Ecol. Lett.* **11**, 1351–1363 (2008).
41. P. L. Zarnetske, D. K. Skelly, M. C. Urban, *Science* **336**, 1516–1518 (2012).
42. C. Riginos, L. M. Porensky, K. E. Veblen, T. P. Young, *Ecol. Appl.* **28**, 323–335 (2018).
43. B. R. Silliman *et al*., *Annu. Rev. Ecol. Evol. Syst.* **44**, 503–538 (2013).
44. T. C. Coverdale *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2101676118 (2021).
45. E. C. Fricke, A. Ordonez, H. S. Rogers, J.-C. Svenning, *Science* **375**, 210–214 (2022).
46. J. L. Atkins *et al*., *Science* **364**, 173–177 (2019).
47. B. R. Silliman, Q. He, *Trends Ecol. Evol.* **33**, 492–503 (2018).

ACKNOWLEDGMENTS

We are grateful for the thoughtful comments provided by E. Bernhardt, T. C. Coverdale, H. Hillebrand, D. Wang, T. P. Young, and five anonymous reviewers on early versions of this manuscript. **Funding:** This work was supported by the National Key Basic Research Program (2022YFC2601100 and 2022YFC3105402), the National Natural Science Foundation of China (31988102 and 32271601), and the Global-ERCaN project of the Chinese Academy of Sciences (177GJH2022020BS). J.S.L. was supported by the Michael E. Tennenbaum Secretarial Scholar gift to the Smithsonian Institution. T.M.P. was supported by the National Science Foundation (DEB-1556905). **Author contributions:** Conceptualization: Q.H., B.R.S.; Methodology: C.X., J.C., X.L., Q.H.; Investigation: C.X., J.C., X.L., Q.H.; Visualization: C.X., Q.H.; Funding acquisition: Q.H., B.L., J.S.L., T.M.P.; Project administration: Q.H.; Supervision: Q.H., B.R.S., B.L.; Writing – original draft: Q.H., C.X. Writing – review and editing: All co-authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available from the sources in the supplementary materials. The Global Consumer Effects dataset, the Global Plant Interactions and Restoration dataset, and code can be assessed at Zenodo (*16*).
License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add2814
Materials and Methods
Figs. S1 to S19
Tables S1 to S15
References (*48*–*686*)
MDAR Reproducibility Checklist

Submitted 1 June 2022; resubmitted 23 March 2023
Accepted 21 September 2023
10.1126/science.add2814

MPOX

APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016

Áine O'Toole[1]*, Richard A. Neher[2], Nnaemeka Ndodo[3], Vitor Borges[4], Ben Gannon[5], João Paulo Gomes[4,6], Natalie Groves[7], David J. King[8], Daniel Maloney[1], Philippe Lemey[9], Kuiama Lewandowski[5], Nicholas Loman[7,10], Richard Myers[7], Ifeanyi F. Omah[1,11], Marc A. Suchard[12], Michael Worobey[13], Meera Chand[7,14], Chikwe Ihekweazu[3], David Ulaeto[7]†, Ifedayo Adetifa[3]†, Andrew Rambaut[1]*†

Historically, mpox has been characterized as an endemic zoonotic disease that transmits through contact with the reservoir rodent host in West and Central Africa. However, in May 2022, human cases of mpox were detected spreading internationally beyond countries with known endemic reservoirs. When the first cases from 2022 were sequenced, they shared 42 nucleotide differences from the closest mpox virus (MPXV) previously sampled. Nearly all these mutations are characteristic of the action of APOBEC3 deaminases, host enzymes with antiviral function. Assuming APOBEC3 editing is characteristic of human MPXV infection, we developed a dual-process phylogenetic molecular clock that—inferring a rate of ~6 APOBEC3 mutations per year—estimates that MPXV has been circulating in humans since 2016. These observations of sustained MPXV transmission present a fundamental shift to the perceived paradigm of MPXV epidemiology as a zoonosis and highlight the need for revising public health messaging around MPXV as well as outbreak management and control.

Since 2017, the Nigeria Centre for Disease Control has been reporting cases of MPXV (mpox virus) infection in humans (fig. S1) (*1*). MPXV, a DNA virus in the genus *Orthopoxvirus*; Family *Poxviridae*, is often described as being endemic in West and Central Africa as a zoonotic disease that transmits through contact with the rodent reservoir host. Since the first human cases were observed in the 1970s, MPXV infections have been predominantly associated with infants and children (*2*–*4*). However, of the cases observed in Nigeria since 2017, very few have been in children, with the virus mainly affecting adults aged 20 to 50 (79%), 27% of which were in women (*5*). Genome sequencing of viruses revealed enough genetic diversity among cases that distinct zoonotic events were not ruled out. In May 2022, cases of MPXV infection were detected spreading widely across Europe and subsequently across the globe. The first MPXV genome sequences from these 2022 cases showed that they had descended from the clade characterized by cases diagnosed in Nigeria and Israel, Singapore, and the UK (*6*) associated with travel from Nigeria (fig. S2 and table S1, in bold). These early 2022 genomes are indicated as a triangle within clade IIb in Fig. 1A and represent lineage B.1 as per the nomenclature proposed by Happi *et al.* (*7*). Isidro *et al.* (*6*) noticed that sequences within lineage B.1 shared 42 single-nucleotide differences from the closest earlier MPXV genomes from 2018. From a 2017 outbreak of MPXV in chimpanzees, the evolutionary rate of MPXV was estimated to be 1.9×10^{-6} substitutions per site per year (1.2×10^{-6} to 2.7×10^{-6}), corresponding to ~1 nucleotide change every 3 years (*8*). Forty-two substitutions in the space of 3 to 4 years is an unexpectedly large number.

Under the paradigm that MPXV is a zoonotic virus with limited human-to-human transmission, one interpretation of this long branch might be that it represents adaptation to humans, facilitating the sustained transmission that is now observed. However, as we show here, and as was quickly seen when the first genomes from 2022 were sequenced, it is clear that these mutations are not the result of errors by the virus's replication machinery and occur at a much higher rate than would be expected for an orthopoxvirus (*6*, *9*). Specifically, most of observed nucleotide changes appear to be of a particular type—a dinucleotide change from TC→TT or its reverse complement, GA→AA (*9*, *10*). This particular mutation is characteristic of the action of the APOBEC3 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3) family of cytosine deaminases. These act on single-stranded DNA (ssDNA) to deaminate cytosine to uracil, causing a G→A mutation in the complementary strand when it is synthesized. Most human APOBEC3 molecules have a strong bias toward deaminating 5′TC dinucleotides, and APOBEC3-driven deamination has been demonstrated with many DNA viruses and retroviruses; (*11*–*18*). Furthermore, a recent study has specifically demonstrated APOBEC3F editing in cell culture and during human MPXV infection (*19*).

We assess the extent to which APOBEC3 has acted on MPXV and explore whether this is the source of the elevated mutation rate observed since 2017. We also explore the evolutionary consequences of this mechanism driving evolution in MPXV and model the distinct processes underpinning the evolution of MPXV in the human population.

APOBEC3 editing as a signature of MPXV evolution in the human population

The known diversity of MPXV is decomposed into three major clades: clades I, IIa, and IIb (Fig. 1A) (*7*). Clade I represents MPXV sampled in Central Africa, and clade IIa is composed of viruses from human and nonhuman animal samples taken in or connected to West Africa. Both of these clades include virus genomes spanning from the 1970s to the present day, although most samples were collected within the last 20 years (fig. S3). Clade IIb has an early sample taken in 1971 (GenBank accession KJ642617), but most of the sequences in clade IIb are more recent virus genomes from 2017 to 2022 that Happi *et al.* (*7*) have labeled as hMPXV-1 (Fig. 1A, labeled phylogeny in fig. S2). Within the recent diversity of clade IIb (indicated as the darker box within IIb in Fig. 1A), we cataloged transmitted mutations that occurred between 2017 and 2022 with only a single representative from the 2022 global lineage B.1 (n = 44 genome sequences, fig. S2) (*20*).

Within MPXV clade IIb, we observe rates of molecular evolution far greater than that expected for double-stranded DNA viruses and indeed that observed in clades I and IIa of MPXV (*8*) and see this excess accumulation of mutation in samples as early as 2017. The great majority of these mutations are of the type G→A or C→T (90.8%) (Fig. 1B), regardless of sample host species (fig. S4). Comparing MPXV clade IIb with clade I and IIa emphasizes that this pattern is not seen outside of clade IIb (Fig. 1B, labeled phylogenies in fig. S5), nor is it seen when looking at reconstructed mutations within a phylogeny of 46 variola virus (VARV) genomes, the human virus responsible for smallpox (fig. S6). For the other MPXV clades, APOBEC-type mutations are observed at between 8 and 13% frequency, which fits with the expected proportion under standard models of nucleotide

[1]Institute of Ecology and Evolution, University of Edinburgh, Edinburgh EH9 3FL, UK. [2]Biozentrum, University of Basel and Swiss Institute of Bioinformatics, Basel, Switzerland. [3]Nigeria Centers for Disease Control and Prevention, Abuja, Nigeria. [4]National Institute of Health Doutor Ricardo Jorge (INSA), Lisbon, Portugal. [5]UK Health Security Agency, Porton Down, Salisbury SP4 0JG, UK. [6]Veterinary and Animal Research Centre (CECAV), Faculty of Veterinary Medicine, Lusófona University, Lisbon, Portugal. [7]UK Health Security Agency, London E14 5EA, UK. [8]CBR Division, Defence Science and Technology Laboratory, Salisbury SP4 0JQ, UK. [9]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium. [10]University of Birmingham, Birmingham B15 2TT, UK. [11]Department of Parasitology and Entomology, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria. [12]Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA. [13]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85719, USA. [14]UKHSA Guys and St Thomas' NHS Trust, London SE1 7EH, UK.
***Corresponding author. Email: aine.otoole@ed.ac.uk (Á.OT.); a.rambaut@ed.ac.uk (A.R.)**
†These authors contributed equally to this work.

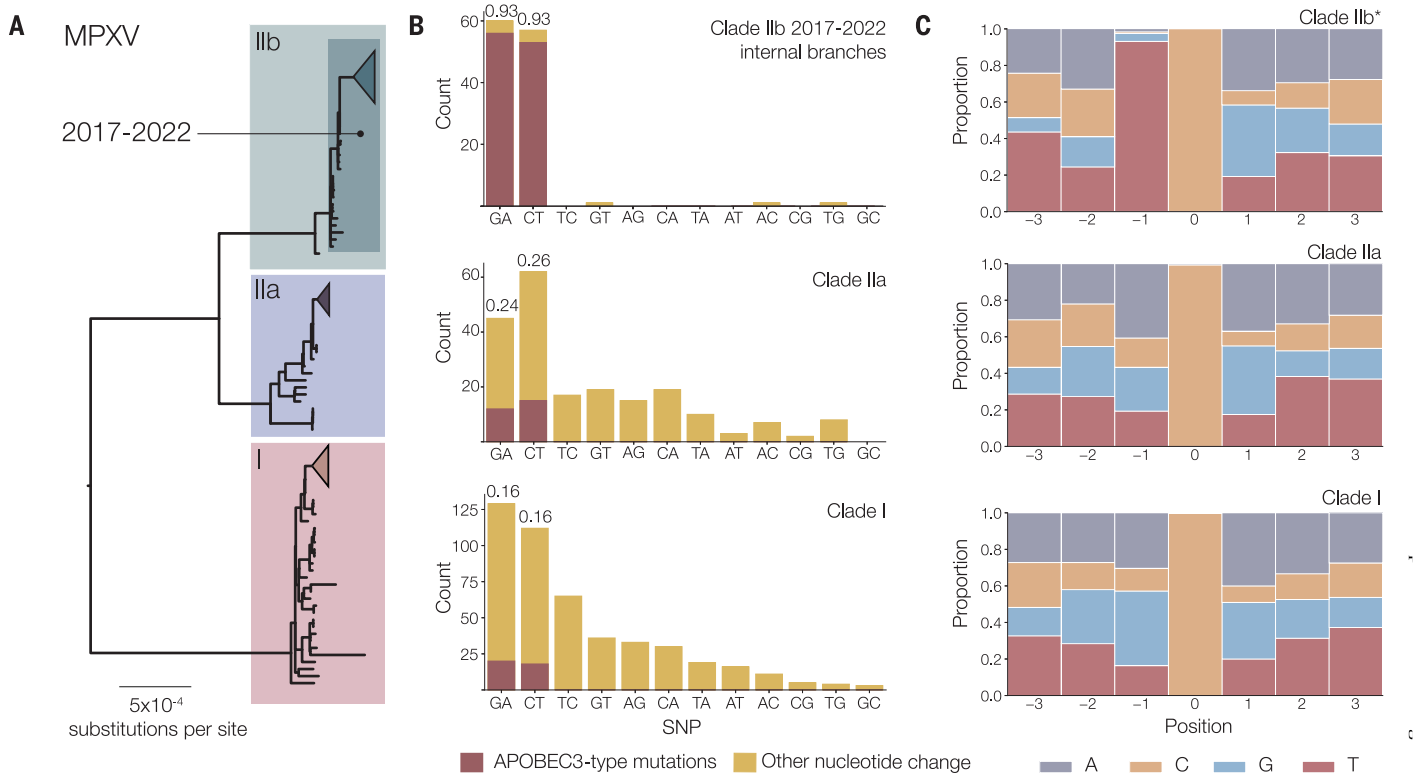


Fig. 1. Specific enrichment of APOBEC3-type mutations in MPXV samples collected since 2017. (**A**) MPXV genetic diversity is categorized into clade I (predominantly sequences from the Democratic Republic of the Congo), clade IIa (predominantly West African sequences), and clade IIb. Within clade IIb is a subclade of genomes sampled from 2017 to 2022 that show mutational patterns distinct from those of the other two clades. (**B**) We catalog single-nucleotide mutations across the phylogenies of clade IIb, clade IIa, and clade I (top to bottom). For clade IIb, we include samples from 2017 to 2022 and only a single representative of the global lineage B.1. Of 120 reconstructed mutations that occurred on internal branches of the clade IIb phylogeny (so are observed transmitted mutations), 109 are consistent with APOBEC3 editing (90.8% of mutations). Individual proportions of G→A and C→T mutations are shown above the respective bars. Ancestral state reconstruction performed across clade IIa and clade I does not produce the same enrichment of mutations consistent with APOBEC3 editing, with only 27 of 207 observed mutations (13%) and 38 of 463 clade I mutations (8%) fitting the dinucleotide pattern. (**C**) Observed heptamers of C→T or G→A mutated sites of clade IIb, IIa, and I phylogenies (top to bottom). Heptamers associated with G→A mutations have been reverse-complemented to reflect deamination on the negative strand. For clade IIb, most C→T mutations are present in a TC dimer context, consistent with APOBEC3 editing (107 of 115 mutations, or 93%). However, the same is not seen for clades IIa and I, in which 29 of 149 (19%) mutations and 42 of 256 (16%) mutations have the dinucleotide context of APOBEC3, respectively, which is what we would predict under standard models of nucleotide evolution. *Only mutations occurring on internal branches of the clade IIb phylogeny are included.

evolution (*21*, *22*) (Fig. 1B). Notably, the heptamers of C→T and G→A mutations that occurred across the clade IIb phylogeny show that this is a specific enrichment of APOBEC3-type dimer mutations of the type TC→TT and GA→AA (Fig. 1C). Similarly, this enrichment of TC→TT and GA→AA mutations is observed within the B.1 lineage (fig. S7), where 84.8% of observed single-nucleotide polymorphisms are consistent with APOBEC3 editing (fig. S8). Observed heptamers around the observed C→T and G→A mutations show a pronounced enrichment in TC and GA target sites in the genomes sampled from 2017 to 2022, in contrast with the rest of MPXV diversity (Fig. 1C), and this enrichment is also reflected within lineage B.1 (fig. S9).

Our analysis highlights that evolution within clade IIb before the emergence of lineage B.1 mirrors that within lineage B.1 but is distinct from MPXV clade I or IIa. Since 2022, the B.1 lineage has been sampled and sequenced internationally in the global epidemic of MPXV. Lineage B.1 is known to be circulating by sustained human-to-human transmission and as such, mutations that have accumulated in B.1 can be considered characteristic of this mode of transmission. We suggest that the APOBEC3-driven evolution of recent clade IIb MPXV is a signature of a switch to sustained transmission within the human population. Within the B.1 lineage, believed to be entirely the result of human infection and transmission, we continue to see the same pattern of predominantly APOBEC3 mutations accumulating at a rate similar to that seen in A lineage genomes since 2017. It is unlikely that, by chance, MPXV evolved to become susceptible to APOBEC3 action within the putative rodent reservoir before the emergence of cases and to retain that susceptibility to human APOBEC3 molecules once transmitting in humans. Given that all human cases sequenced since 2017 share substantial numbers of APOBEC3 mutations, including nine on the stem branch leading to hMPXV-1, it is very unlikely that these represent multiple zoonotic introductions. A*POBEC3* genes emerged in placental mammals from a duplication of the ancestral *AID* gene and have a dynamic recent evolutionary past, with gene duplication and loss across phyla (*23*, *24*). *APOBEC3* genes in primates have undergone recent expansion, with primate genomes now having seven paralogs of three ancestral genes (*25*, *26*). Rodents, the reservoir of MPXV, have only a single functional APOBEC3 protein, resulting from gene loss and fusion events (*25*). Rodent *APOBEC3* has been shown to be expressed preferentially in spleen and bone marrow, with limited expression observed in other tissues (*27*, *28*).

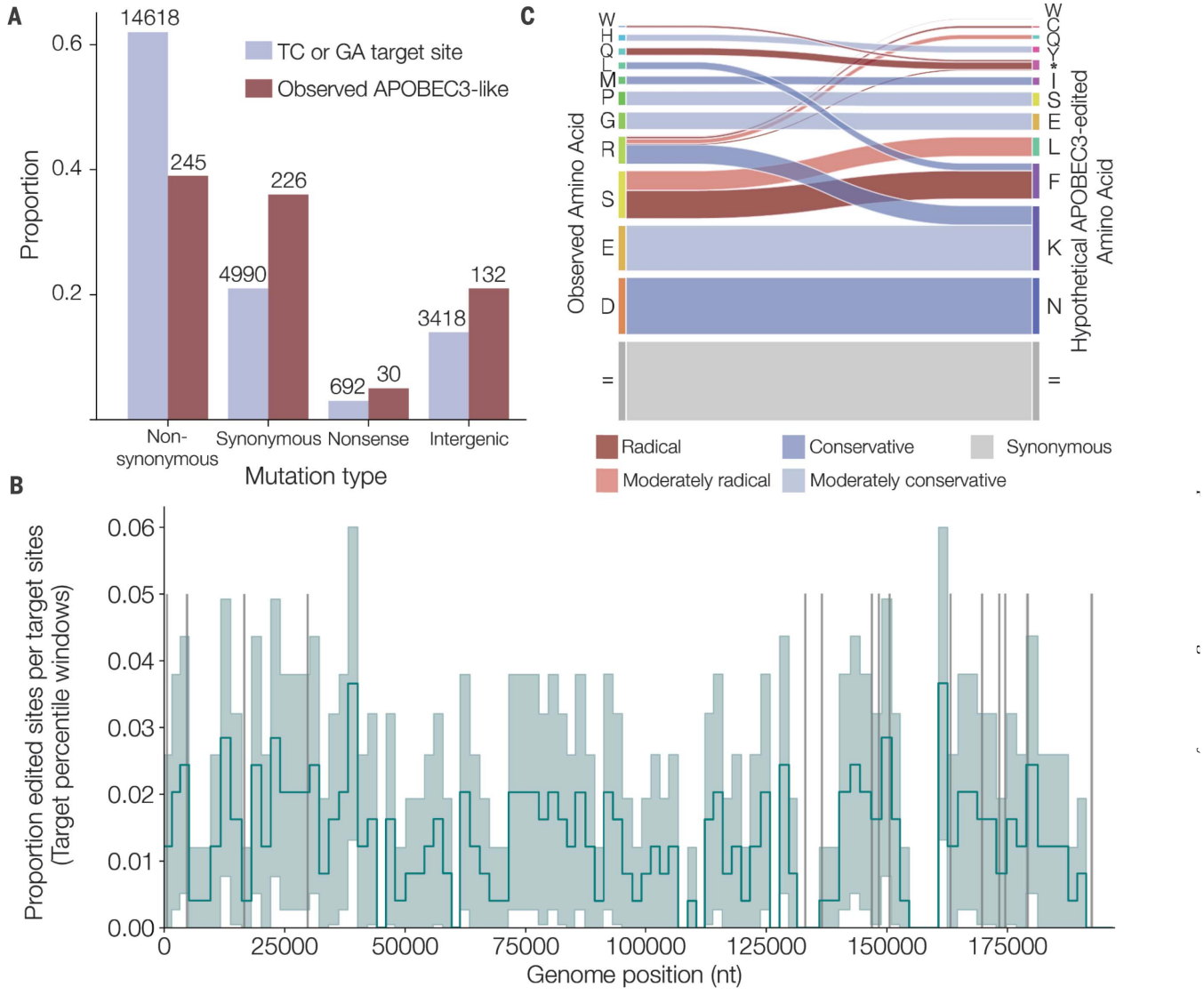


Fig. 2. Observed APOBEC3-type mutations are not merely a product of available target sites. (**A**) Consequence of hypothetical APOBEC3 mutations at target dimer site (either the C in the TC target site or G in the GA target site) in the coding regions of the National Center for Biotechnology Information reference MPXV genome for clade II (accession NC_063383) and those observed APOBEC3 mutations across the coding regions of the clade IIb phylogeny (not including the outgroup branch leading to the 1971 genome sequence). These are categorized into nonsynonymous (altered amino acid), synonymous (amino acid remaining unchanged), nonsense (editing producing a stop codon), and intergenic (not present in a coding sequence). (**B**) The proportion of target sites edited for each target site percentile window across the MPXV genome. The teal shaded regions represent the binomial confidence interval around observations. Masked regions are indicated by vertical gray bars. Observed edits include data from the clade IIb phylogeny, with a single representative of lineage B.1 and not including the branch leading to the outgroup 1971 genome sequence. (**C**) Hypothetical amino acid changes for codons overlapping with TC and GA target sites in a reference MPXV genome (GenBank accession number: NC_063383) if APOBEC3 edited those dimers to TT and AA. Amino acid changes are colored by Grantham Score (0–50 conservative, dark blue; 51–100 moderately conservative, light blue; 101–150 moderately radical, light red; >150 radical, dark red; synonymous, gray). Single-letter abbreviations for the amino acid residues are as follows: C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; W, Trp; and Y, Tyr. Stop codons are indicated with an asterisk (*).

APOBEC3 has a limited repertoire to generate variation in the MPXV genome

If we assume this observed evolution within hMPXV-1 is APOBEC3-driven, this may have implications for its sustained transmission in the human population. Considering all GA and TC dimer sites in the clade II reference genome (accession number NC_063383)—i.e., those that could be the target of APOBEC3 editing but had not been by that point—we assessed what amino acid changes a deamination mutation at these sites would bring about (Fig. 2). Of the 23,718 such dimers, 61.6% (14,618) would produce amino acid replacements, 21% (4990) would be synonymous, 2.9% (692) would induce stop codons, and 14.4% (3418) would occur outside of coding regions. For the clade IIb genomes, of the 633 mutations at these dimers that did occur, 38.7% (245) were amino acid replacements and 35.7% (226) were synonymous, 4.7% (30) were nonsense, and a further 132 APOBEC3 mutations were in intergenic regions (20.8%). The probability of getting 226 or greater synonymous mutations out of 663 under a simple binomial distribution with 21.0% chance of a context being synonymous is $P = 7.6 \times 10^{-18}$. We do not see the same enrichment for synonymous mutations in the mutations that are not APOBEC3-like, although the quantity of these mutations is considerably lower (fig. S10).

There are also more mutations outside of protein-coding regions than we would expect based on the location of target dimers (probability of 4.5×10^{-6} of getting at least 132 noncoding mutations given that only 14.4% of targets are in these regions). This supports the hypothesis that what we observe are the residual least harmful APOBEC3 mutations after natural selection has eliminated those with substantial fitness costs to the virus. By comparing the density of observed C→T and G→A mutations and the density of APOBEC3 target sites (TC or GA dinucleotides) across the reference genome, we see that the distribution of mutations is not simply a product of the availability

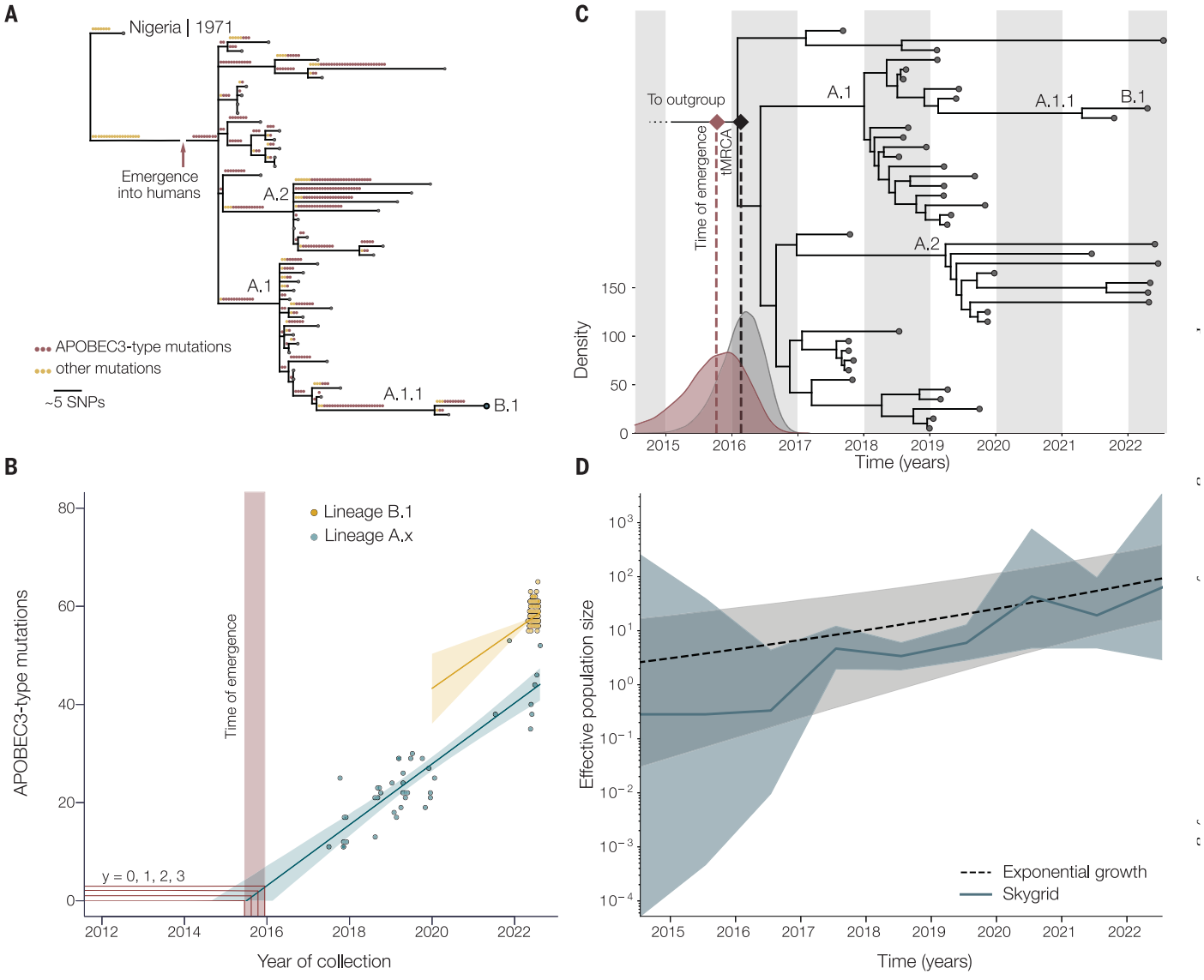


Fig. 3. Estimating the time of MPXV emergence into the human population from the accumulation of APOBEC3-type mutations. (**A**) MPXV genomes sampled from human infections from 2017 to 2022, with an outgroup sequence from an outbreak in Nigeria in 1971 (n = 44 genome sequences, including outgroup). Lineages indicated as per nomenclature proposed by Happi *et al*. (*7*). Mutations along each branch are indicated with circles colored by whether it is putatively APOBEC3 edited (TC→TT and GA→AA; red) or whether it is another mutation type (yellow). The break in the basal branch illustrates the assumption made in the regression model in panel B, that all APOBEC3 mutations occurred after emerging into the human population; however, we do not know the precise distribution of red or yellow mutations. (**B**) APOBEC3 mutations from MPXV genomes sampled since 2017. The reconstructed most recent common ancestor (MRCA) of the panel A phylogeny is used as the root in the root-to-tip plot, and the y intercept is used as a proxy for time of emergence, which is inferred by fitting a Bayesian regression to the sequence dates from panel A. Intersects with y = 1, 2, and 3 are also shown as it is likely that a small number of the APOBEC3-type mutations are actually earlier replication errors and not induced by APOBEC3. (**C**) Maximum clade credibility (MCC) phylogeny of MPXV clade IIb with absolute time shown on the x axis. We separated the alignment into an APOBEC3 and a non-APOBEC3 partition and modeled the substitution process in each independently. We used an epoch model with two outgroup sequences (not shown) representing the first epoch and hMPXV-1 ingroup sequences representing MPXV post-emergence into the human population with an exponential growth model. The probability density distributions show the estimated time of the most recent common ancestor (tMRCA) of the ingroup as well as the estimated transition time that represents the time of emergence into the human population. (**D**) Estimated effective population size of the outbreak using a nonparametric coalescent Skygrid model with 11 change points over a period of 8.5 years. This reconstruction falls within the bounds of the exponential growth model estimated from the second epoch in panel C, suggesting that the MPXV population has been exponentially growing since at least 2016.

of target sites (Kolmogorov-Smirnov test statistic = 0.07, $P = 0.0002$; fig. S11, A and B). When considering synonymous and nonsynonymous APOBEC3-like mutations separately, there is a significant difference between the density of target sites across the MPXV genome and the distribution of observed APOBEC3-like synonymous and nonsynonymous mutations, respectively (fig. S11, C and D).

The "repertoire" of mutations that APOBEC3 is able to provide as genetic variation on which natural selection can act is severely restricted. Only a limited number of dinucleotide contexts are present, and the repertoire of amino acid changes that APOBEC3 editing can induce is also limited (Fig. 2C and fig. S12). Only 13 different amino acid replacements are possible, and three that give rise to stop codons, and they are not reversible by the same mechanism. This means that given the restricted set of positions at which these mutations occur and the limited amino acid changes that they can result in, the elevated rate will not necessarily facilitate adaptation of the virus.

APOBEC3 hypermutation is a host-mediated antiviral mechanism. These molecules act as the viral genome is being replicated and single strands are exposed. During repeated rounds of replication, either strand can be deaminated, leading to both C→T and G→A changes on the positive strand, as seen in this study. Thus, it is likely that the genomes that are extensively mutated by APOBEC3 will simply not be viable and will not be transmitted further. MPXV replicates in the cytoplasm likely by means of rolling-circle amplification (*29*), and this facilitates extensive continuous genome replication (*30*). The high processivity of this mechanism efficiently produces high copy numbers of MPXV genome molecules in the cell, potentially saturating the APOBEC3 enzyme action if the concentration of MPXV DNA molecules is high enough. This likely means that many MPXV genomes are unaffected by APOBEC3 action. Occasionally, however, a genome, modestly mutated by APOBEC3, may remain viable and be transmitted. We see this in the enrichment of observed synonymous and intergenic mutations relative to available targets in the MPXV genome. Given the nonreversibility of the APOBEC3 action, sustained evolution within the human population may result in a depletion of lower-consequence target sites (i.e., synonymous or conservative amino acid changes) and thus expedite a decrease in fitness of MPXV over time. This could be both through a reduction in the number of viable offspring viruses produced by infected cells and as a result of the accumulation of moderately deleterious mutations by genetic drift (i.e., mutation load). However, the timescale on which this might happen is uncertain and other evolutionary forces such as recombination may act to restore fitness, and we do not address this further in this study. A further uncertainty arises when considering the variable repertoire of genes associated with virus infectivity or host immune modulation in poxvirus genomes. Mutations that alter or abrogate the function of these genes may have little direct effect on virus replication machinery but may disrupt the virus–host interaction. There is precedent for the naturally occurring inactivation of genes in VARV contributing to host specificity, and consequently the loss of function of some MPXV genes through APOBEC3 activity may potentially have adaptive value for the virus as it replicates and transmits in a new host (*31*, *32*).

Even if the mutations that accumulate through this process are simply the neutral residue of a suboptimal antiviral host defense, they have produced sufficient variability for the phylogenetic analysis of the epidemic over the short term. The initial lineages proposed by Happi *et al.* (*7*) have expanded with the 2022 epidemic B.1 lineage now encompassing 17 sublineages at the time of writing (*33*). The rapid and temporally linear accumulation of mutations means that genomic epidemiological models and tools (*34*, *35*), usually used for RNA viruses, may also have utility for hMPXV-1.

The linear accumulation of APOBEC3-type mutations since the emergence and spread of MPXV in humans

Since 2017, the genomes thus far sampled from clade IIb have accumulated APOBEC3-type single-nucleotide mutations approximately linearly over time (Fig. 3, A and B; labeled phylogeny in fig. S13). We applied Bayesian regression analysis on the root-to-tip plot of sequences in Fig. 3A, which includes one representative B.1 genome, and also separately on the B.1 lineage (B.1 phylogeny in fig. S1) (*20*). To ensure that the elevated temporal signal is specific to APOBEC3 data, we show combinations of APOBEC3 and non-APOBEC3 mutations on clade IIb data in fig. S14. The estimated rate of accumulation was 6.18 per year (95% credible intervals of 5.20, 7.16). For the B.1 lineage, the rate was 5.93 per year (2.95, 8.92), suggesting that despite widespread and rapid transmission within MSM (men who have sex with men) networks, the rate of accumulation of APOBEC3 mutations remained the same as it did for the rest of clade IIb. It is notable that the regression line for B.1 lies substantially above that for the rest of clade IIb, suggesting that this lineage accumulated more mutations than expected before the emergence of B.1. However most of these mutations are also present in the genome from Maryland, USA (accession number ON676708) from November 2021 (*10*), indicating that they arose and circulated for some months before the B.1 epidemic (figs. S15 and S16). Extrapolating back to when the APOBEC3-type mutations started to accumulate provides an estimate of when the first APOBEC3 mutations occurred in the stem of the branch leading to the 2017 epidemic. If we assume that all these mutations are due to APOBEC3 in humans, then we estimate this date of emergence to be 20 June 2015. However, we expect a few APOBEC3-like mutations to actually be due to replication errors during the earlier preemergence epoch. The number of mutations that we ascribe to this period will affect our estimate linearly—i.e., if three mutations were not due to APOBEC3, then the estimate would shift to 14 December 2015.

To accommodate this uncertainty in our estimates, we have developed a more explicit model of APOBEC3-mediated evolution in the BEAST software package (*20*, *34*). We estimate that the action of APOBEC3 on the MPXV population is driving evolution ~28 times faster than the background evolutionary rate (fig. S17). Gigante *et al.* (*10*) also described an elevated overall rate of evolution in the A lineage but did not decompose the APOBEC3 and non-APOBEC3 contribution to this. The time of the most recent common ancestor of the post 2017 genomes is estimated to be 23 February 2016 (28 June 2015, 28 September 2016), with the transition to sustained human-to-human transmission estimated to be 14 September 2015 (21 August 2014, 31 July 2016; Fig. 3C and fig. S17). Unlike the assumption in Fig. 3B that all APOBEC3 mutations occurred after emergence, the BEAST analysis estimates the transition point integrating over all possibilities. This allows for the fact that a few of the APOBEC3-like mutations may actually be due to replication errors in the earlier evolutionary epoch and this might explain the slightly more recent date. We also see evidence of exponential growth in the number of infections in the epidemic before the emergence of lineage B.1 in 2022 (Fig. 3D), despite the decline in cases reported in 2020 (fig. S1C), albeit the growth rate is relatively slow, which reinforces the indication from the demographics of the cases that this is not a generalized epidemic.

Implications for the global public health response to mpox cases

Since the identification of the B.1 lineage, a number of countries have reported other lineages that lie outside the diversity of B.1, including the United States, United Kingdom, Portugal, India, and Thailand. In almost all instances, these cases are reported as having a history of international travel. The lineages in which these are placed (designated as A.2.1, A.2.2, A.2.3, and A.3) can all be phylogenetically traced back to the epidemic in Nigeria (Fig. 3A). This suggests that at least one instance of sustained human-to-human transmission is still ongoing outside of the recognized MSM networks that were the focus of the 2022 global epidemic. Stopping transmission in these communities, though necessary, will not be sufficient to eliminate the virus as a human epidemic. Many countries lack the surveillance

to detect MPXV cases, and if sustained human-to-human transmission has been ongoing since 2015–2016, it is plausible that there are other populations that are currently enduring epidemics.

Historically, mpox was considered a zoonotic disease, and cases have been treated as independent spillover events with low levels of circulation in the human population. Thus far, this continues to be an accurate characterization of clade I in Central Africa. For clade IIb, although some non-B.1 lineage cases may be new zoonotic infections, most cases since 2016 are likely the result of human-to-human transmission. Although the B.1 lineage across the world is now diminished—though not yet eradicated—the human epidemic from which it arose continues unabated. It is critical that global public health affords MPXV cases in countries that are historically considered to have endemic reservoir species equal attention and concern to those elsewhere. Surveillance needs to be global if MPXV is to be eliminated from the human population and then prevented from reemerging.

REFERENCES AND NOTES

1. NCDC, "An Update of Monkeypox Outbreak in Nigeria" (Nigeria Centre for Disease Control, 2017), (available at https://ncdc.gov.ng/diseases/sitreps/?cat=8).
2. A. W. Rimoin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16262–16267 (2010).
3. T. D. Baker, in *Clio Medica. Acta Academiae Internationalis Historiae Medicinae* (Brill, 1982), vol. 17, pp. 268–269.
4. Z. Jezek *et al*., *J. Trop. Med. Hyg.* **90**, 31–38 (1987).
5. E. Alakunle, U. Moens, G. Nchinda, M. I. Okeke, *Viruses* **12**, 1257 (2020).
6. J. Isidro *et al*., *Nat. Med.* **28**, 1569–1572 (2022).
7. C. Happi *et al*., *PLOS Biol.* **20**, e3001769 (2022).
8. L. V. Patrono *et al*., *Nat. Microbiol.* **5**, 955–965 (2020).
9. Á. O'Toole, A Rambaut, Initial observations about putative APOBEC3 deaminase editing driving short-term evolution of MPXV since 2017. *Virological* (2022); https://virological.org/t/830.
10. C. M. Gigante *et al*., *Science* **378**, 560–565 (2022).
11. R. Suspène, M. Henry, S. Guillot, S. Wain-Hobson, J.-P. Vartanian, *J. Gen. Virol.* **86**, 125–129 (2005).
12. H. Chen *et al*., *Curr. Biol.* **16**, 480–485 (2006).
13. J.-P. Vartanian, D. Guétard, M. Henry, S. Wain-Hobson, *Science* **320**, 230–233 (2008).
14. J.-P. Vartanian *et al*., *PLOS Pathog.* **6**, e1000928 (2010).
15. M. Tsuge *et al*., *Virus Res.* **149**, 211–216 (2010).
16. R. Suspène *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4858–4863 (2011).
17. Z.-G. Peng *et al*., *Hepatology* **53**, 1080–1089 (2011).
18. M. Fehrholz *et al*., *J. Gen. Virol.* **93**, 565–576 (2012).
19. R. Suspène *et al*., *J. Infect. Dis.*jiad165 (2023).
20. See materials and methods.
21. M. Hasegawa, H. Kishino, T. Yano, *J. Mol. Evol.* **22**, 160–174 (1985).
22. M. Kimura, *J. Mol. Evol.* **16**, 111–120 (1980).
23. S. G. Conticello, C. J. F. Thomas, S. K. Petersen-Mahrt, M. S. Neuberger, *Mol. Biol. Evol.* **22**, 367–377 (2005).
24. A. Krishnan, L. M. Iyer, S. J. Holland, T. Boehm, L. Aravind, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3201–E3210 (2018).
25. C. Münk, A. Willemsen, I. G. Bravo, *BMC Evol. Biol.* **12**, 71 (2012).
26. L. Yang, M. Emerman, H. S. Malik, R. N. McLaughlin Jnr., *eLife* **9**, e58436 (2020).
27. M. C. Mikl *et al*., *Mol. Cell. Biol.* **25**, 7270–7277 (2005).
28. A. I. Su *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4465–4470 (2002).
29. B. Moss, *Cold Spring Harb. Perspect. Biol.* **5**, a010199 (2013).
30. Q. Peng *et al*., *Science* **379**, 100–105 (2023).
31. R. C. Hendrickson, C. Wang, E. L. Hatcher, E. J. Lefkowitz, *Viruses* **2**, 1933–1967 (2010).
32. B. Mühlemann *et al*., *Science* **369**, eaaw8977 (2020).
33. GitHub, mpxv-lineages/lineage-designation: Official place for proposals and details around monkeypox virus lineage designations; https://github.com/mpxv-lineages/lineage-designation.
34. M. A. Suchard *et al*., *Virus Evol.* **4**, vey016 (2018).
35. P. Sagulenko, V. Puller, R. A. Neher, *Virus Evol.* **4**, vex042 (2018).
36. A. Rambaut, Á. O'Toole, hmpxv/apobec3, Zenodo (2023); https://doi.org/10.5281/zenodo.8234483.

ACKNOWLEDGMENTS

We thank those involved in sequencing MPXV genome sequences and shared data online rapidly, facilitating rapid public health responses, and also those who have continued to contribute to open science by sharing their data in open access platforms such as GenBank. We have provided all data and scripts related to this manuscript on Zenodo (*36*). **Funding:** Wellcome Trust ARTIC (Collaborators Award 206298/Z/17/Z, ARTIC network) (Á.O.T., P.L., M.A.S., A.R.); European Research Council (grant agreement no. 725422 – ReservoirDOCS) (P.L., M.A.S., A.R.); National Institutes of Health (R01 AI153044) (P.L., M.A.S., A.R.); David and Lucile Packard Foundation (M.W.); Research Foundation, Flanders–Fonds voor Wetenschappelijk Onderzoek–Vlaanderen, G066215N, G0D5117N and G0B9317N (P.L.); HORIZON 2020 EU grant 874850 MOOD (P.L.); HERA project (grant/2021/PHF/23776) supported by the European Commission through the European Centre for Disease Control and Prevention (V.B. and J.P.G.). Crown Copyright © 2023. The Nigeria Centre for Disease Control and Prevention receives core funding from the Nigerian government. **Author contributions:** Conceptualization: Á.O.T., A.R.; Methodology: Á.O.T., R.N., A.R.; Investigation: N.N., V.B., B.G., J.P.G., N.G., D.K., D.M., K.L., R.M., I.F.O., M.W.; Visualization: Á.O.T., R.N., A.R.; Funding acquisition: A.R., P.L., M.S., M.C., C.I., D.U., I.A.; Supervision: D.U., I.A., A.R.; Writing–original draft: Á.O.T., R.N., P.L., M.S., M.W., A.R.; Writing–review and editing: Á.O.T., R.N., N.N., V.B., B.G., J.P.G., N.G., D.K., D.M., P.L., K.L., N.L., R.M., I.F.O., M.S., M.W., M.C., C.I., D.U., I.A., A.R.
Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** Data, code, and materials used in the analysis are available from Zenodo (*36*). Tables S1 and S2 include details of accession numbers and author lists for all source genome sequence data used. All except a single genome sequence were sourced from GenBank on NCBI. Data partitions of the alignments have been included in the XML files on Zenodo and GitHub for ease of reproduction. A single sequence was sourced from GISAID (gisaid.org) and as such was removed from the XMLs. Supplementary tables include the GISAID identifier, which can be used to independently source the data, and instructions for constructing the alignment and addition to the XML are included. **License information:** Copyright © 2023 the authors, some rights reserved, including, where applicable, UK Crown copyright licensed under the Open Government License v.3.0; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. In the interest of rapid dissemination of results with immediate public health relevance, and because this research was funded in whole or in part by the Wellcome Trust (Collaborators Award 206298/Z/17/Z, ARTIC network), a cOAlition S organization, the author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg8116
Materials and Methods
Figs. S1 to S17
Tables S1 to S4
References (*37*–*40*)
MDAR Reproducibility Checklist

Submitted 23 January 2023; resubmitted 1 June 2023
Accepted 21 September 2023
10.1126/science.adg8116

DEVELOPMENT

The –*KTS* splice variant of WT1 is essential for ovarian determination in mice

Elodie P. Gregoire[1], Marie-Cécile De Cian[1]†, Roberta Migale[2]†, Aitana Perea-Gomez[1], Sébastien Schaub[3], Natividad Bellido-Carreras[1], Isabelle Stévant[4,5], Chloé Mayère[4,5], Yasmine Neirijnck[1], Agnès Loubat[1], Paul Rivaud[6], Miriam Llorian Sopena[2], Simon Lachambre[7], Margot M. Linssen[8], Peter Hohenstein[8], Robin Lovell-Badge[2], Serge Nef[4,5], Frédéric Chalmel[6], Andreas Schedl[1], Marie-Christine Chaboissier[1]*

Sex determination in mammals depends on the differentiation of the supporting lineage of the gonads into Sertoli or pregranulosa cells that govern testis and ovary development, respectively. Although the Y-linked testis-determining gene *Sry* has been identified, the ovarian-determining factor remains unknown. In this study, we identified –KTS, a major, alternatively spliced isoform of the Wilms tumor suppressor WT1, as a key determinant of female sex determination. Loss of –*KTS* variants blocked gonadal differentiation in mice, whereas increased expression, as found in Frasier syndrome, induced precocious differentiation of ovaries independently of their genetic sex. In XY embryos, this antagonized *Sry* expression, resulting in male-to-female sex reversal. Our results identify –KTS as an ovarian-determining factor and demonstrate that its time of activation is critical in gonadal sex differentiation.

In mice, sex is genetically determined by the constitution of the sex chromosomes. This leads to testis or ovary development in XY and XX embryos, respectively, which, in turn, influences the sexual development of the whole individual. Before sex determination, WNT/β-catenin signaling mediated by R-spondin1 (RSPO1) contributes to the proliferation of the gonadal progenitors in both sexes (*1*). In XY gonads, at around embryonic day (E) 11.5, RSPO1/WNT/β-catenin is down-regulated and *Sry* and its direct target *Sox9* are up-regulated in a subset of progenitors derived from the overlying coelomic epithelium (*2*–*4*). These transcription factors induce Sertoli cell differentiation. Once differentiated, they no longer express *Sry* but express other genes, including *Amh* (*5*–*7*), and establish the fate of the testis. In XX gonads, pregranulosa cell differentiation occurs slightly later, around E12.0 to 12.5, as shown by their loss of bipotentiality (*8*–*10*), the de novo expression of the transcription factor FOXL2 (*11*), and the stabilization of RSPO1/WNT/β-catenin signaling (*12*, *13*). However, the gene(s) initiating ovarian differentiation have remained unknown (*14*).

One of the key factors in the early development of the gonad is the Wilms tumor suppressor WT1, a zinc-finger transcriptional regulator (*15*). *WT1* (human)/*Wt1* (mouse) encodes two major alternative spliced isoforms that do or do not include the three amino acids KTS (lysine, threonine, and serine) between the two last zinc fingers. These isoforms are named +KTS and –KTS, respectively. Whereas –KTS acts as a transcriptional activator or repressor depending on the cellular context, the insertion of +KTS abrogates DNA binding and promotes the subnuclear localization of WT1 in nuclear speckles (*16*, *17*). A simple imbalance of the ratio of both isoforms in favor of –*KTS* is the molecular basis of Frasier syndrome, characterized by male-to-female sex reversal (*18*, *19*) associated with the down-regulation of *Sry* as evidenced in the mouse model (*20*).

Results

Distribution of –KTS transcripts during gonadal development

To determine the distribution of WT1 splice variants in E11.5 XY mouse gonads, we carried out BaseScope in situ hybridizations. Scoring revealed cellular heterogeneity, with cells containing variable amounts of +*KTS* or –*KTS* transcripts (Fig. 1A and fig. S1A). This observation was confirmed with single-cell RNA sequencing (scRNA-seq) analysis of the splice junction reads obtained from sorted cells dissected from E11.5 mouse gonads (Fig. 1B). Next, we examined single-cell transcriptomic data of the supporting cell lineage in both sexes from E10.5 to E13.5 (*8*) (Fig. 1C and fig. S1B). Although +*KTS* exhibited similar mRNA levels between XY and XX gonads at E10.5 and E[illegible] –*KTS* transcripts were detected in greater amounts in XY gonads at E11.5 before increasing in XX gonads at E12.5, time points that coincide respectively with Sertoli and pregranulosa cell differentiation.

–KTS is required for the differentiation of the supporting cells

To address the contribution of –*KTS* to sex determination, we revisited the mouse model of –*KTS* ablation (–*KTS*$^{-}$/–*KTS*$^{-}$ is denoted as –*KTS KO*) that results in gonadal dysgenesis (*20*) (fig. S2, A and B). We performed single-cell transcriptome profiling of wild-type and mutant gonadal cells collected at around E12.0 (Fig. 1, D to I, and table S1). Cells were projected in a two-dimensional (2D) space by using uniform manifold approximation and projection (UMAP) and partitioned into 39 clusters (Fig. 1D). Cluster annotation identified Sertoli and pregranulosa cells in the controls determined from the expression of known markers (Fig. 1, E to G, and fig. S3, A and B); however, these clusters were not present in –*KTS KO* gonads (Fig. 1, E to H, and fig. S3B). Nevertheless, presupporting cells were observed in –*KTS* mutants of both sexes, as revealed by the expression of *Runx1* (mRNA)/RUNX1 (protein) at E12.0 [fig. S3B, cluster 3 (c3)] and at E12.5 (Fig. 2A and fig. S4). XY –*KTS*-deficient gonads exhibited a few scattered cells expressing SOX9, contrasting with the widespread SOX9-positive Sertoli cells that formed nascent testis cords in XY controls (Fig. 2B). Furthermore, XY –*KTS*-deficient gonads were devoid of *Amh*/AMH expression (Fig. 2, C and D, and fig. S5, A and B) and instead abnormally maintained *Rspo1* and SRY at E12.5 and until birth (fig. S2, C and E, and fig. S5, C and E). Together, our results indicate that the presupporting cells did not differentiate as bona fide Sertoli cells in the absence of –KTS. Despite the significantly reduced expression of *Sox9* (P value = 0.0035), the XY –*KTS KO* presupporting cells failed to differentiate into pregranulosa cells, as demonstrated by the almost complete absence of *Foxl2*/FOXL2 expression at E12.5 (Fig. 2, B and D, and fig. S3). Similarly, *Foxl2*/FOXL2 expression was strongly reduced in XX –*KTS KO* gonads, further supporting the importance of –KTS for the differentiation of the pregranulosa lineage and the activation of the female program (Fig. 2, B and D, and fig. S3). At around birth, the expression of *Sox9*/SOX9 and *Foxl2*/FOXL2 remained low in –*KTS*-deficient gonads of both sexes, and SOX9/FOXL2 double-positive cells were detected (fig. S5, F to H), indicating poor differentiation of the supporting lineage. Together, these data demonstrate that –*KTS* is dispensable for the specification of presupporting cells but is necessary to stabilize Sertoli cell differentiation and essential to initiate pregranulosa cell differentiation.

[1]Université Côte d'Azur, INSERM, CNRS, Institut de Biologie Valrose (iBV), 06108 Nice, France. [2]The Francis Crick Institute, London NW1 1AT, UK. [3]Sorbonne Université, CNRS, Development Biology Laboratory (LBDV), 06234 Villefranche sur Mer, France. [4]Department of Genetic Medicine and Development, University of Geneva, 1211 Geneva, Switzerland. [5]iGE3, Institute of Genetics and Genomics of Geneva, University of Geneva, 1211 Geneva, Switzerland. [6]Univ Rennes, INSERM, EHESP, IRSET (Institut de Recherche en Santé, Environnement et Travail)-UMR_S 1085, 35000 Rennes, France. [7]Infinity, INSERM, CNRS, University Toulouse III, 31024 Toulouse, France. [8]Central Animal and Transgenic Facility and Department of Human Genetics, Leiden University Medical Center, 2333ZA Leiden, Netherlands.
***Corresponding author. Email: marie-christine.chaboissier@univ-cotedazur.fr**
†These authors contributed equally to this work.

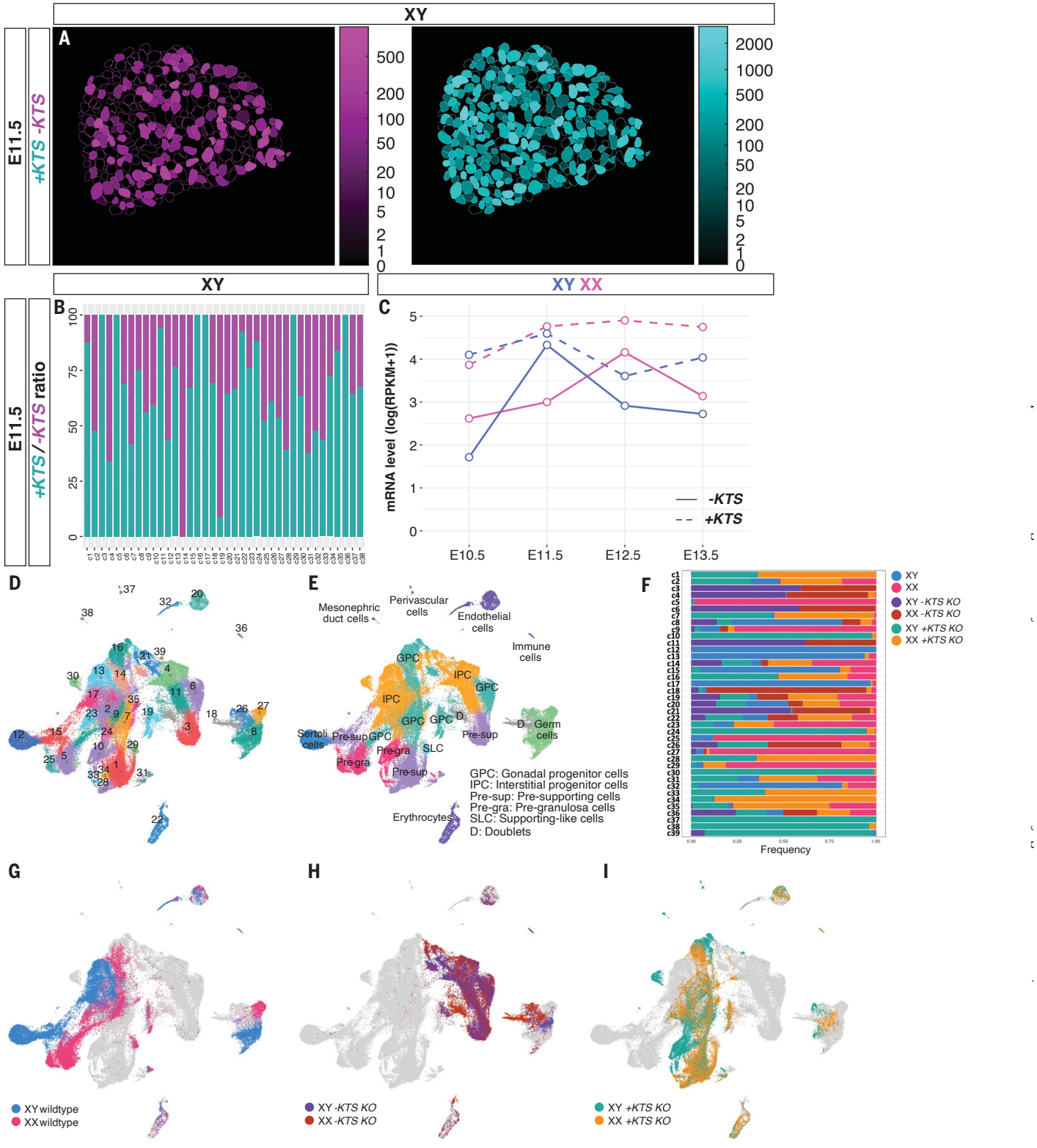


Fig. 1. Dynamic distribution of +KTS and −KTS transcripts and single-cell transcriptomic analysis of −KTS KO and +KTS KO during early mouse-gonad development. (**A**) Representative area distribution of *−KTS* (magenta) and *+KTS* (cyan) transcripts from Basescope in situ hybridizations on XY gonad sections at E11.5 [21 tail somites (ts)] in square micrometers per nucleus measured with DicHysto protocol. Data from both gonads are representative of biological and technical duplicates. (**B**) *−KTS* and *+KTS* mRNA ratio in E11.5 (21 ts) XY wild-type individual cells. (**C**) *+KTS* and *−KTS* transcript levels in single-cell transcriptomic dataset of differentiating supporting cells. RPKM, reads per kilobase million. (**D**) UMAP projection of the 75,360 cells colored by clusters or (**E**) by associated cell types. (**F**) Association of cell clusters with genotypes. (**G** to **I**) UMAP projection by genotypes with (G) XY (blue) and XX wild type (pink); (H) XY (purple) and XX *−KTS KO* (*−KTS*$^{-}$/*−KTS*$^{-}$) (brown); and (I) XY (green) and XX *+KTS KO* (*+KTS*$^{-}$/*+KTS*$^{-}$) gonads (orange).

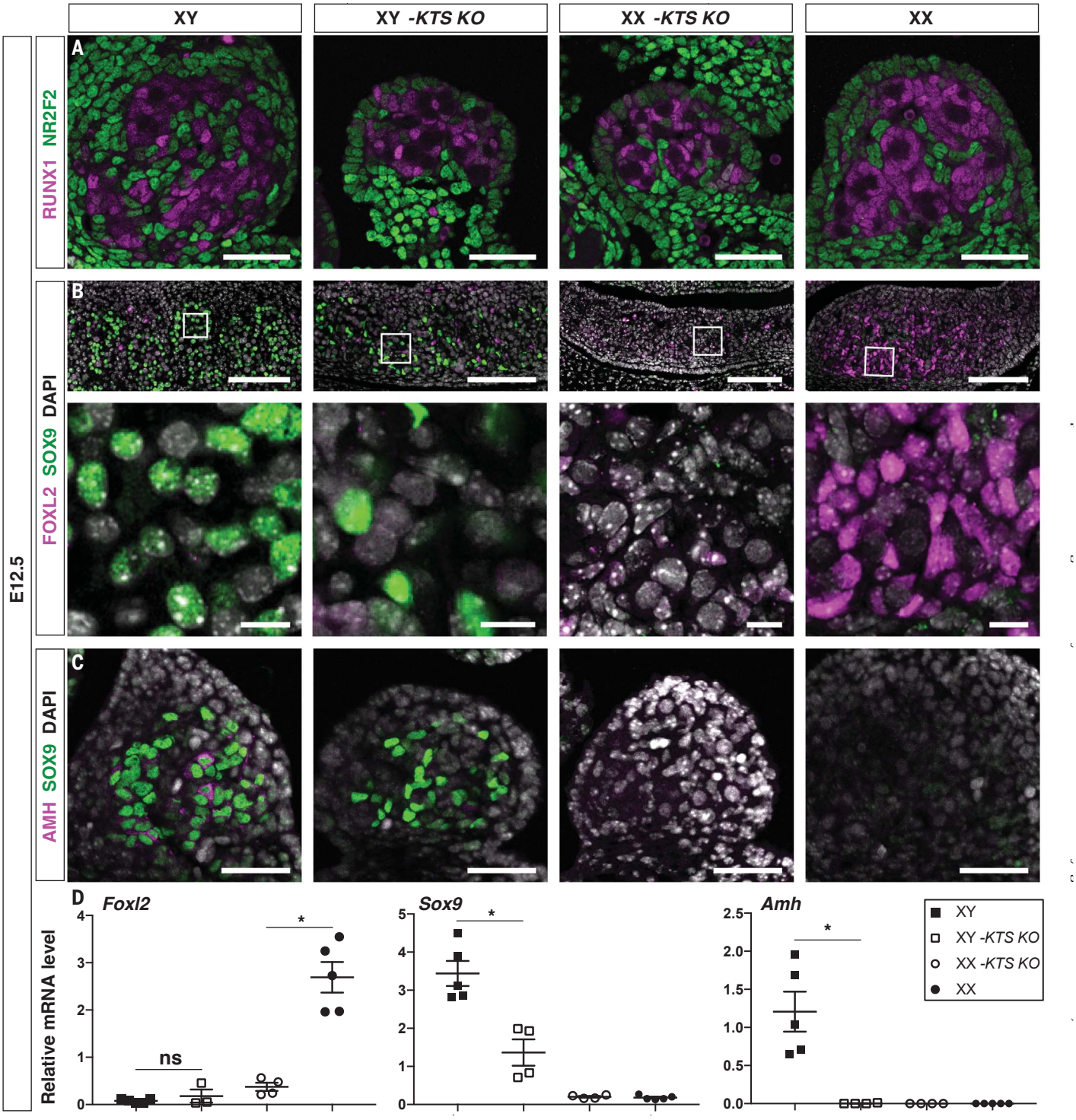


Fig. 2. −KTS is necessary for sex differentiation of the supporting cells. Immunodetection of (**A**) the presupporting cell marker RUNX1 (magenta) and the progenitor marker NR2F2 (green) at E12.5 (scale bars, 50 μm), (**B**) the Sertoli cell marker SOX9 (green) and pregranulosa cell marker FOXL2 (magenta) (scale bars, 100 μm and 10 μm, respectively), and (**C**) SOX9 (green) and AMH (magenta) (scale bars, 50 μm) in the indicated genotypes. Data for (A) to (C) are representative of triplicate biological replicates. Nuclei labeled with 4′,6-diamidino-2-phenylindole (DAPI) are shown in white. (**D**) Quantification of *Foxl2*, *Sox9*, and *Amh* transcripts after normalization to *Gapdh* by RT-qPCR. Data are shown as means ± SEM. *−KTS KO* denotes $-KTS^{-}/-KTS^{-}$.

Absence of +KTS *triggers an increase of* −KTS *amounts*

In patients with Frasier syndrome, heterozygous mutations in WT1 prevent the production of *+KTS,* resulting in higher amounts of *−KTS* variants (*18*, *19*). Given the role of *−KTS* in ovarian determination, we investigated the contribution of this increase to sex reversal in the *+KTS KO* ($+KTS^{-}/+KTS^{-}$) mouse model (*20*). At E12.5, XY *+KTS* mutant gonads were enriched for RUNX1- and FOXL2-positive pregranulosa cells and contained rare Sertoli cells, as expected for male-to-female sex reversal (Fig. 3A and fig. S6A). Next, we verified that *−KTS* transcripts were twice as abundant in

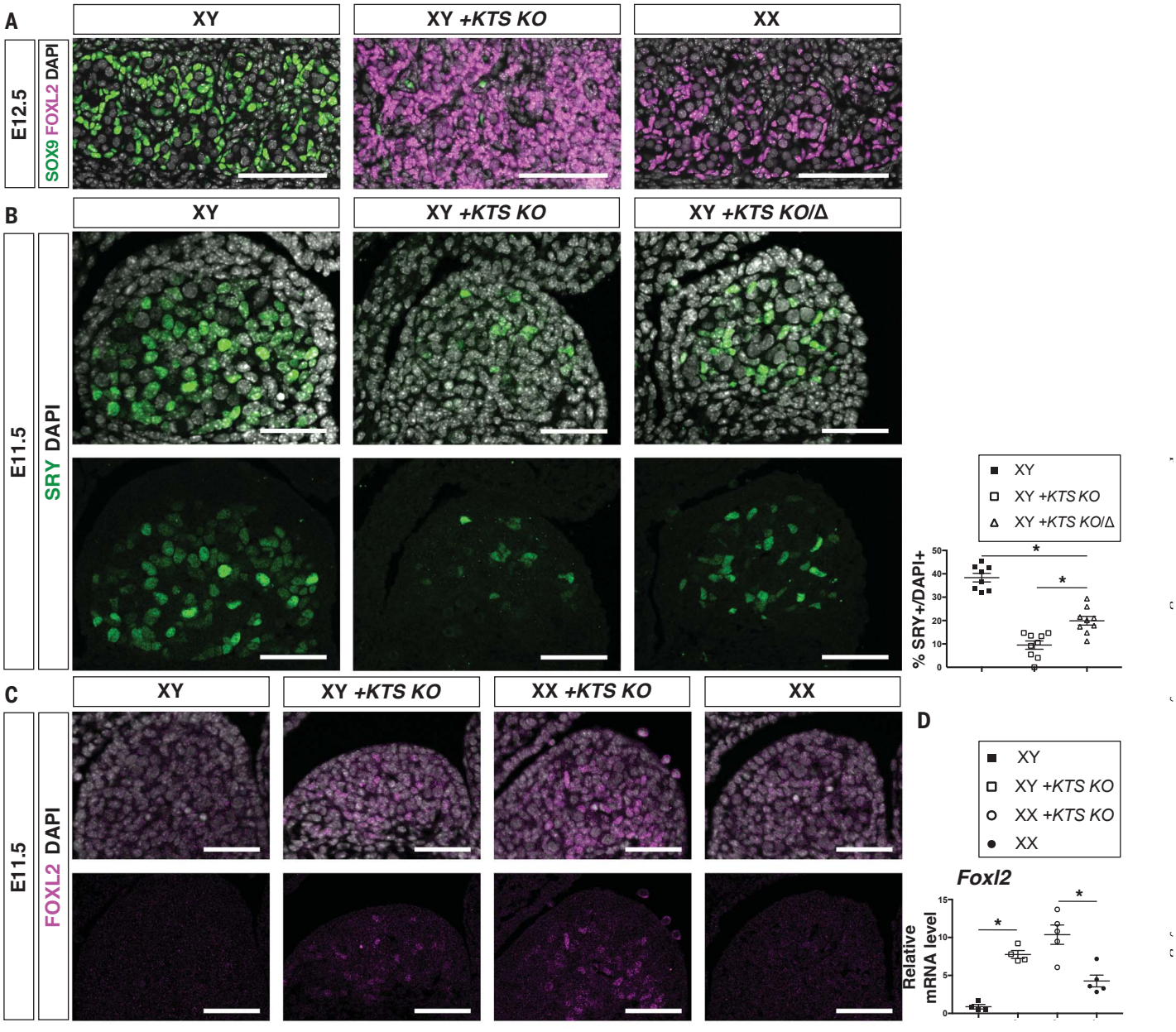


Fig. 3. Early pregranulosa cell differentiation occurs in XY and XX *+KTS KO* gonads. (**A**) Immunofluorescence of the Sertoli cell marker SOX9 (green) and the pregranulosa cell marker FOXL2 (magenta) in indicated genotypes at E12.5 (n = 4 embryos). Scale bars, 100 μm. (**B**) Immunostaining of SRY (green) at E11.5 (21 ± 1 ts) in XY, XY *+KTS KO* (*+KTS*$^{-}$/*+KTS*$^{-}$), and XY compound heterozygotes (*+KTS KO*/Δ denotes *+KTS*$^{-}$/*Wt1*$^{-}$). Scale bars, 50 μm. Quantification of SRY+ cells normalized to DAPI+ cells labeled in white in the upper panel. n = 4 embryos, two sections per embryo. Data are shown as means ± SEM. (**C**) Immunodetection of the pregranulosa cell marker FOXL2 in indicated genotypes of triplicate biological replicates at 20 to 21 ts. Scale bars, 50 μm. (**D**) Relative mRNA expression of *Foxl2* normalized to *Gapdh* at 20 to 21 ts. Data are shown as means ± SEM.

+KTS KO gonads as in controls (fig. S6B). In addition, total WT1 protein levels were similar in XY *+KTS KO* and control gonads, confirming that absence of +KTS is compensated with an increase of −KTS isoforms (fig. S6C).

Precocious pregranulosa cell differentiation prevents Sry activation in the mouse Frasier model

Sex reversal in *+KTS KO* embryos is caused by a failure of *Sry* activation (*20*, *21*). To determine if this arises from the lack of *+KTS* or from an increase in *−KTS* variants, we compared the number of SRY-positive cells in *+KTS KO* and *+KTS KO*/Δ compound embryos, both of which lack alleles encoding *+KTS* and contain two and one allele encoding *−KTS,* respectively. The number of SRY-expressing cells was higher in *+KTS KO*/Δ than in XY *+KTS KO* gonads, indicating that *Sry* expression does not require +KTS but is antagonized by the higher level of −KTS (Fig. 3B). Furthermore, *Rspo1* was abundant in XY *+KTS KO* gonads at E11.5, a stage when it is down-regulated in XY gonads, and *Foxl2*/FOXL2 expression was markedly elevated in XY and XX *+KTS KO* gonads (Fig. 3, C and D, and fig. S6E). This suggests precocious pregranulosa cell differentiation irrespective of the genetic sex.

−KTS is sufficient to induce ovarian development

Single-cell transcriptome profiling of XY and XX *+KTS KO* gonads at E12.0 (Fig. 1, D to F,

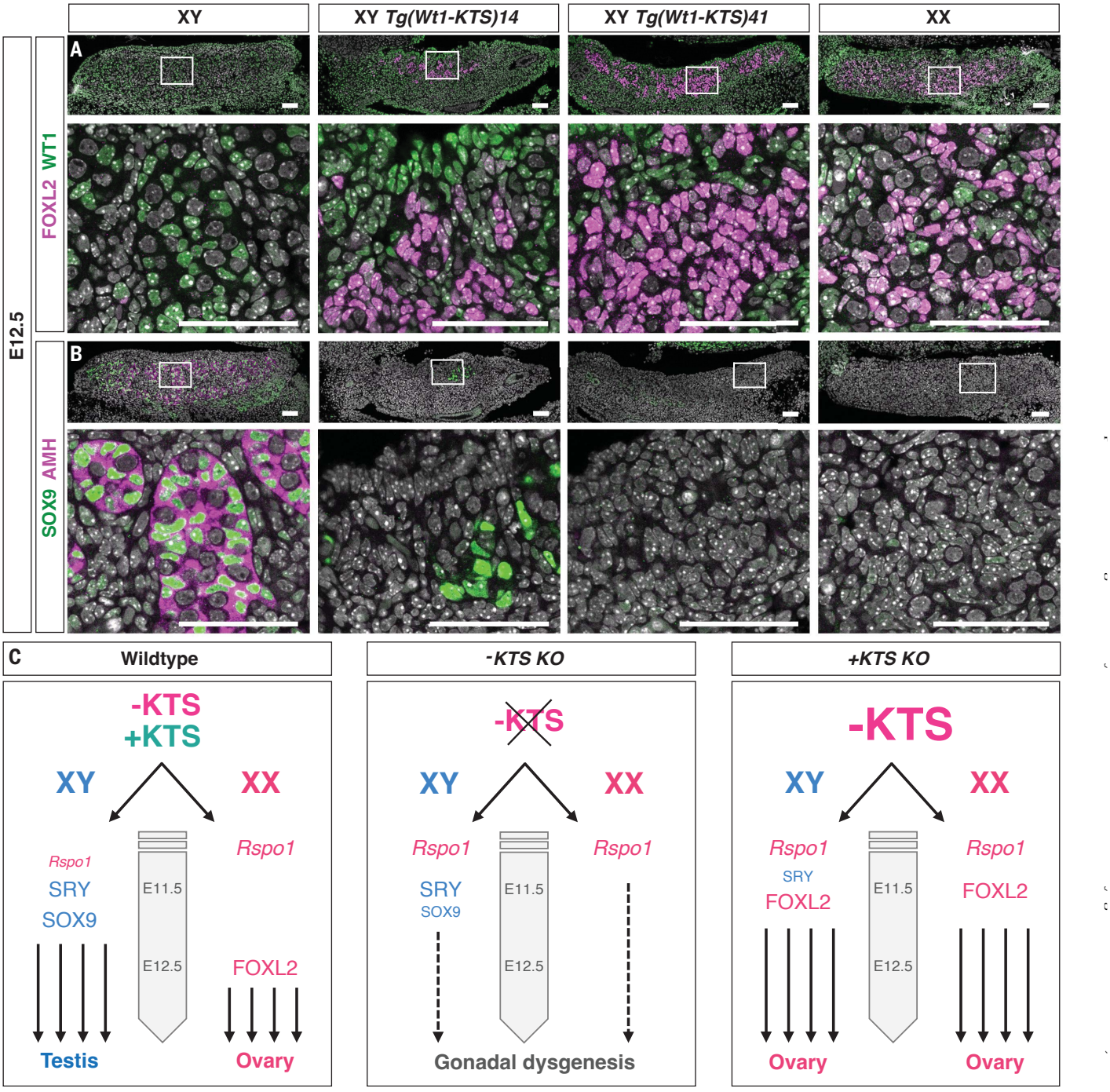


Fig. 4. –KTS induces pregranulosa cell differentiation in XY transgenic gonads. (**A**) Immunofluorescence of the pregranulosa cell marker FOXL2 (magenta) and WT1 (green) in the indicated genotypes at E12.5. Scale bars, 50 μm. (**B**) Immunostaining of the Sertoli cell markers SOX9 (green) and AMH (magenta) in indicated genotypes. Scale bars, 50 μm. (**C**) Model of supporting cell differentiation in wild-type and *KTS* mutant gonads: Absence of *–KTS* in *–KTS KO* gonads promotes the maintenance of *Rspo1* transcripts and impairs SOX9 and FOXL2 expression, leading to gonadal dysgenesis. Increasing –KTS in *+KTS KO* gonads results in ovarian differentiation in both genetic sexes.

and I, and fig. S3B) identified two pregranulosa cell clusters (c10 and c33) distinct from those found in XX controls (c5 and c25) and from XY Sertoli cells (c12). Further comparison of transcriptomes of these clusters confirmed that E12.0 *+KTS KO* cells are transcriptionally related to pregranulosa cells (fig. S7). Next, we used comparative analysis to identify genes that are activated or repressed by –KTS in the context of female sex determination (fig. S8). In XX *–KTS KO* presupporting cells, the expression of 319 genes was significantly deregulated [false discovery rate (FDR)–adjusted *P* value ≤ 0.05, data S4]. *Pdgfa* and *Tcf21,* reported to be targets of WT1 in other organs (*22*, *23*), were down-regulated, whereas *Igf2*—a direct target of WT1 (*24*)—and genes highly expressed in bipotent presupporting cells including *Sprr2d* (*25*), *Wnt6* (*26*), and *Nr0b1* (*27*), were up-regulated in XX *–KTS KO* presupporting cells and down-regulated in XX *+KTS KO*

pregranulosa cells, suggesting that they are repressed by –KTS during sex determination. Altogether, our data suggested that increased –KTS, rather than loss of +KTS, was responsible for XY sex reversal in the *+KTS KO* model. To test whether –KTS was sufficient to induce ovarian differentiation in a XY wild-type gonad, we performed transient additive transgenesis using a bacterial artificial chromosome (BAC) construct covering the *Wt1* locus, in which we introduced the classical Frasier mutation in intron 9 (interference with +KTS production). Three out of four XY transgenic animals showed the presence of FOXL2-positive cells, indicating that –KTS promotes differentiation of pregranulosa cells in XY gonads (Fig. 4, A and B, and fig. S9, A and B). Moreover, reverse transcription quantitative polymerase chain reaction (RT-qPCR) analysis of genotypes producing different levels of –KTS suggested that *–KTS* must reach a threshold to robustly activate *Foxl2* expression (fig. S9C).

We can conclude that the altered expression of *+KTS* caused by mutations in the donor splice site in intron 9 of *Wt1* promotes an increase of the amount of –KTS, which, in turn, prematurely activates ovarian differentiation, prevents *Sry* up-regulation, and impairs testis development (Fig. 4C). –KTS thus represents a key actor in gonad development that is required to initiate ovarian development.

Discussion

Here, we provide evidence that sex determination depends not only on the up-regulation of the sex-determining factors *Sry* and *–KTS* for male and female fates, respectively, but also on their timing (*28*–*30*). This is an important concept because *–KTS* is an autosomal factor expressed in both XY and XX gonads. In wild-type mice, *Sry* acts before *–KTS*, thus securing testis development in XY gonads. If *Sry* expression is impaired or delayed, or if *–KTS* is up-regulated prematurely, such as in the Frasier syndrome model (*+KTS KO*), the pregranulosa cell differentiation is accelerated, resulting in male-to-female sex reversal. After the peak of SRY action (*30*), –KTS becomes necessary to maintain Sertoli cell differentiation in XY embryos and to initiate pregranulosa cell differentiation in XX embryos (Fig. 4C). Although differences in timing and dynamics of sex determination make a direct comparison between mouse and human data difficult, the sex-reversal phenotype in mice carrying intron 9 mutations suggests that this is a good mouse model for the human Frasier syndrome. Our data thus indicate that increased expression of –KTS, rather than loss (or reduction) of +KTS, is the primary cause of sex reversal in Frasier syndrome. Notably, a change of *+KTS*/*–KTS* ratio in favor of *–KTS* operates when the eggs of *Chelydra serpentina,* a turtle with temperature-dependent sex determination, are shifted from a male- to a female-producing temperature (*31*). This outcome suggests that the –KTS isoform of WT1 is also involved in ovarian determination outside of the mammalian class.

REFERENCES AND NOTES

1. A.-A. Chassot *et al*., *Development* **139**, 4461–4472 (2012).
2. R. Sekido, I. Bar, V. Narváez, G. Penny, R. Lovell-Badge, *Dev. Biol.* **274**, 271–279 (2004).
3. N. Gonen *et al*., *Science* **360**, 1469–1473 (2018).
4. J. Karl, B. Capel, *Dev. Biol.* **203**, 323–333 (1998).
5. M.-C. Chaboissier *et al*., *Development* **131**, 1891–1901 (2004).
6. P. De Santa Barbara *et al*., *Mol. Cell. Biol.* **18**, 6653–6665 (1998).
7. M. Bullejos, P. Koopman, *Dev. Dyn.* **221**, 201–205 (2001).
8. I. Stévant *et al*., *Cell Rep.* **26**, 3272–3283.e3 (2019).
9. K. Harikae *et al*., *J. Cell Sci.* **126**, 2834–2844 (2013).
10. W. Niu, A. C. Spradling, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20015–20026 (2020).
11. D. Schmidt *et al*., *Development* **131**, 933–942 (2004).
12. P. Parma *et al*., *Nat. Genet.* **38**, 1304–1309 (2006).
13. D. M. Maatouk, L. Mork, A.-A. Chassot, M.-C. Chaboissier, B. Capel, *Dev. Biol.* **383**, 295–306 (2013).
14. M. Elzaiat, A.-L. Todeschini, S. Caburet, R. A. Veitia, *Clin. Genet.* **91**, 173–182 (2017).
15. J. A. Kreidberg *et al*., *Cell* **74**, 679–691 (1993).
16. C. Englert *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11960–11964 (1995).
17. S. H. Larsson *et al*., *Cell* **81**, 391–401 (1995).
18. S. Barbaux *et al*., *Nat. Genet.* **17**, 467–470 (1997).
19. B. Klamt *et al*., *Hum. Mol. Genet.* **7**, 709–714 (1998).
20. A. Hammes *et al*., *Cell* **106**, 319–329 (2001).
21. S. T. Bradford *et al*., *Hum. Mol. Genet.* **18**, 3429–3438 (2009).
22. Z. Y. Wang, S. L. Madden, T. F. Deuel, F. J. Rauscher 3rd, *J. Biol. Chem.* **267**, 21999–22002 (1992).
23. R. Bandiera *et al*., *Dev. Cell* **27**, 5–18 (2013).
24. A. Ward *et al*., *Gene* **167**, 239–243 (1995).
25. G. J. Bouma, Q. J. Hudson, L. L. Washburn, E. M. Eicher, *Biol. Reprod.* **82**, 380–389 (2010).
26. A. T. Cory, A. Boyer, N. Pilon, J. G. Lussier, D. W. Silversides, *Mol. Reprod. Dev.* **74**, 1491–1504 (2007).
27. A. Swain, E. Zanaria, A. Hacker, R. Lovell-Badge, G. Camerino, *Nat. Genet.* **12**, 404–409 (1996).
28. E. M. Eicher, L. L. Washburn, *J. Exp. Zool.* **228**, 297–304 (1983).
29. M. Bullejos, P. Koopman, *Dev. Biol.* **278**, 473–481 (2005).
30. R. Hiramatsu *et al*., *Development* **136**, 129–138 (2009).
31. T. Rhen, R. Fagerlie, A. Schroeder, D. A. Crossley 2nd, J. W. Lang, *Differentiation* **89**, 31–41 (2015).
32. S. Schaub, SebastienSchaub/DICHisto: Gregoire_et_al2023, Zenodo (2023); https://doi.org/10.5281/ZENODO.8329139.
33. S. Lachambre, Nuclei count in gonad, Version 3, Zenodo (2023); https://doi.org/10.5281/ZENODO.8345079.
34. S. Lachambre, Intensity measurement in gonad, Version 2, Zenodo (2023); https://doi.org/10.5281/ZENODO.8345091.

ACKNOWLEDGMENTS

We thank M. Cerciat for her help in the preparation of 10X Genomics libraries. Sequencing was performed by the GenomEast platform, a member of the France Génomique consortium (ANR-10-INBS-0009). We are grateful to F. Jian Motamedi for her advice on transgenesis. The microscopy was performed at the MICA facility, at the iBV (Institut de Biologie Valrose, Université Côte d'Azur, CNRS, INSERM, iBV, France). We acknowledge help from members of the PRISM Imaging Platform, from the Animal house at Ibv, and from C. Passot and M. Cutajar-Bossert for the management of funding. We are grateful to the FLOW, BABS, ASF, and BRF facilities of the Francis Crick Institute. We are also grateful to members of the A.S. and M.-C.C. groups for helpful discussions. **Funding:** This research was supported by Agence Nationale de la Recherche grant ANR-19-CE14-0022 SexDiff (M.-C.C.), Fondation Maladies Rares grant FONDATION-GenOmics 2018-1 181003 RNAseq (M.-C.C.), Agence Nationale de la Recherche grant ANR-20-CE14-0022 Ardigerm (M.-C.C.), and Swiss National Science Foundation grants 31003A_173070 (S.N.) and 310030_200316 (S.N.). The Francis Crick Institute receives its core funding from Cancer Research UK (FC001107), the UK Medical Research Council (FC001107), Wellcome (FC001107), and the UK Medical Research Council (U117512772) (R.M., M.L.S., and R.L.-B.). **Author contributions:** E.P.G., R.L.-B., A.S., and M.-C.C. contributed to the conception and design of the study. E.P.G., M.-C.D.C., R.M., A.P.-G., N.B.-C., and A.L. carried out mouse work, in situ hybridization (ISH), immunofluorescence, microscopy, scRNA-seq, and other experimental analysis. S.S. and S.L. performed image posttreatments for quantitative analysis of Basescope ISH and immunostainings. I.S., Y.N., C.M., P.R., M.L.S., and F.C. performed the bioinformatics analysis. E.P.G., M.M.L., P.H., and A.S. generated transgenesis. E.P.G., A.S., and M.-C.C. wrote the first draft of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version. Funding and resources: S.N., R.L.-B., and M.-C.C. Supervision: M.-C.C. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The scRNA-seq datasets are available at the National Center for Biotechnology (NCBI) Gene Expression Omnibus (GEO): GSE207097, and in articles cited in the paper. Programs developed are available at Zenodo (*32*–*34*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by the Swiss National Science Foundation (grants 31003A_173070 and 310030_200316), a cOAlition S organization. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add8831
Materials and Methods
Figs. S1 to S9
Tables S1 to S3
References (*35*–*49*)
Data S1 to S4
MDAR Reproducibility Checklist

Submitted 13 July 2022; resubmitted 3 August 2023
Accepted 29 September 2023
10.1126/science.add8831

WORKING LIFE

By **Shanta Hejmadi**

Staying rooted

I wasn't surprised to learn I hadn't been selected for a prestigious postdoc fellowship. I had applied mostly hoping for feedback that would improve my application the following year. But comments from two reviewers made me rethink my plan to apply again—and even whether to continue in academia. Under the section titled "weaknesses," one stated, "The work would be conducted at the same institution as their Ph.D." Another wrote: "No justification is given for staying at the same institution." The comments were infuriating. I couldn't understand why I (or anyone!) should have to justify not wanting to move for a job that often doesn't support a single person, let alone a family.

I was older than most of my peers when I started my Ph.D., and more settled. I have now lived in Minnesota for 22 years, having moved to the area fresh out of college. My spouse has a good job here that supports us. We own a home and our gender-nonconforming teenager feels welcome and safe in our community. Our extended families also live nearby.

My parents moved to Minnesota specifically to help us with child care. As I pursued my education, support from both sets of grandparents and other members of our extended families was critical. They provided bus pickups, after school care, and generally served as the "village" that it takes to raise a child. With their help, my spouse and I could focus on our careers without worrying about child care. As our parents age, we hope to offer them the same kind of daily support. It's not a situation I can easily leave.

"The reality is that moving is a huge barrier for many scientists."

When I started my Ph.D., I knew that if I wanted to pursue an academic career, I would likely have to move someday. Over the course of my program, I considered long and hard what I—and the rest of my family—would need to sacrifice in order to do that, and the list was long. In the end, I decided I needed to find a way to stay in Minnesota.

So, as I wrapped up my Ph.D., I started to apply for anything local that fit my skills and interests. This included academic postdocs, as well as biologist and data scientist positions in industry and the public sector. One academic opportunity that caught my eye was a postdoc fellowship offered by the National Science Foundation that's designed to broaden participation of groups underrepresented in biology. As a first-generation American with a multiracial background, I thought I'd be a good candidate. So, I reached out to a faculty member at my institution whose research focused on what I wanted to study next. She was interested in my ideas and we developed a proposal.

The instructions clearly stated that applicants who proposed to remain in their current location had to justify their choice. I interpreted that to mean the research needed to be distinctly different from my dissertation, so I made sure to emphasize in the proposal that I'd be working in a new lab on a completely different project.

I didn't mention my other reasons for wanting to stay in Minnesota because it wasn't the reviewers' business; normally, employers aren't allowed to ask about spouses or children. So it came as a shock when they discounted my application because I hadn't disclosed all of my reasons for proposing to stay put.

The whole episode left me wondering what purpose it serves to penalize those who want to stay at the same institution for multiple career stages. It's often said in academia that "moving around a lot is necessary." But the reality is that moving is a huge barrier for many scientists, including those with spouses, children, aging parents, disabilities, and cultural backgrounds that emphasize a close-knit extended family community—the kinds of people academia is actively trying to recruit to its ranks.

Thankfully, my story had a happy ending: After receiving multiple job offers, I opted to take a position as a senior data scientist in a public transit agency, where I work on challenging, interesting problems that have a real impact in the region where I live. I didn't have to move, and I can use skills I developed in graduate school.

For the sake of others, though, I wish funding agencies were more open to supporting scientists who want to stay rooted to one location. What's so essential about moving? ■

Shanta Hejmadi is a senior data scientist at the Metro Transit Division of the Metropolitan Council in Minneapolis and St. Paul, Minnesota.